Pivotal | IT大咖说 知识分享平台

GPDB在百度外卖的实践与平台化

——黄冕

2017.08

百度外卖研发中心

大纲

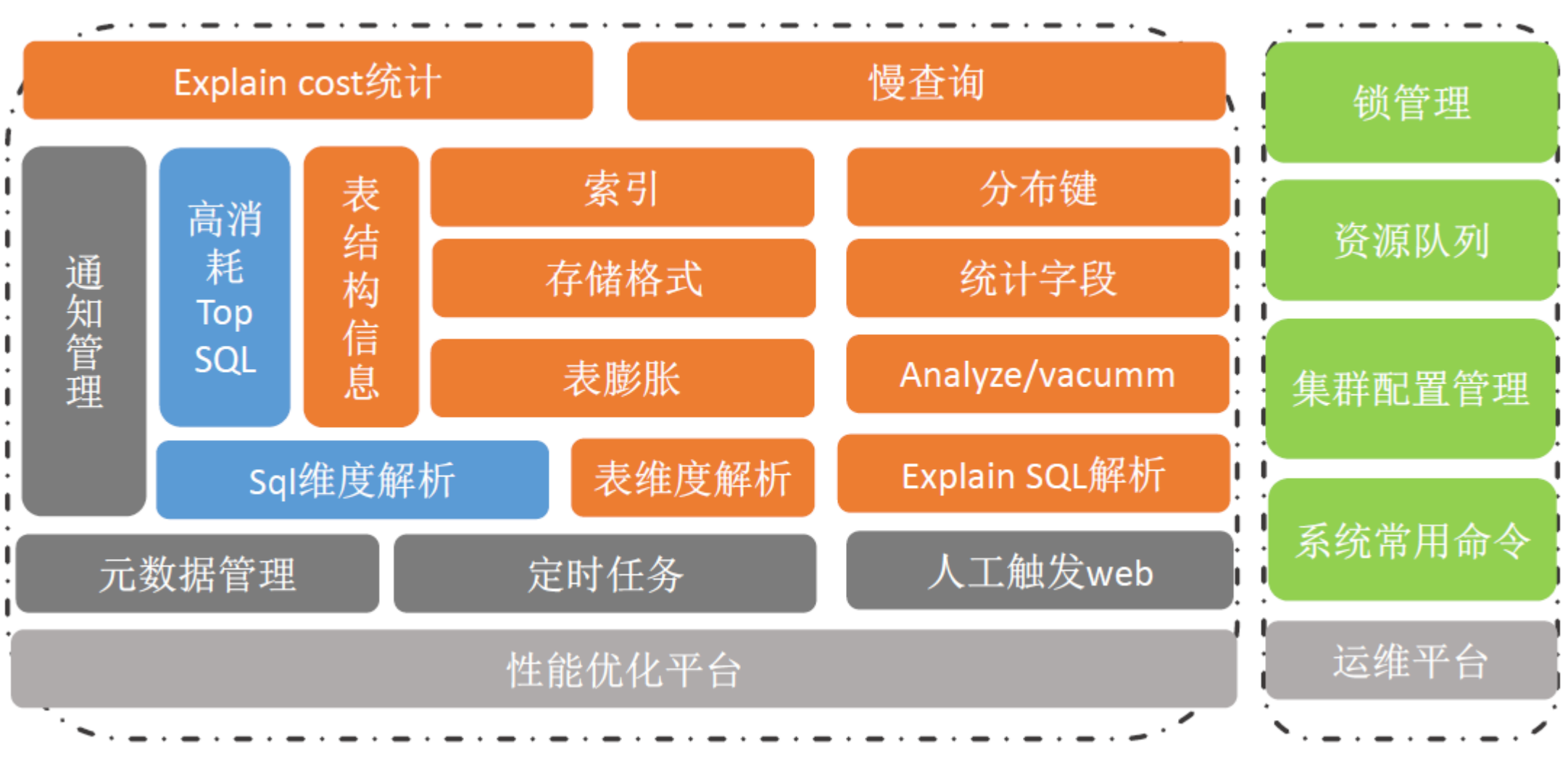
- 为什么平台化
- 分布键管理与选择
- 资源队列管理与动态变更
- 锁管理与深入分析
- 围绕sql分析的优化
- 一些例行运维操作
- 未来与挑战

- 我们的场景
  - 一段时间内多维度大并发查询的需求
  - 一定的实时场景
  - 共有388张表
  - 机器资源机器有限

- 出现的问题
  - 对查询时长有着较为严苛的要求
  - 表数量较大，表数据分布等分析工作量大
  - 在消费的时间内，需要进行生产，导致锁问题
  - 需要合理安排资源，否则资源完全不够用

- 便于快速通过图表发现问题
- 避免重复劳动，提高生产力
- 便于积累，避免零散的脚本
- 提高自身的能力和成就感
- 提高开放性，使得用户参与到sql优化中来

性能优化与运维平台架构图

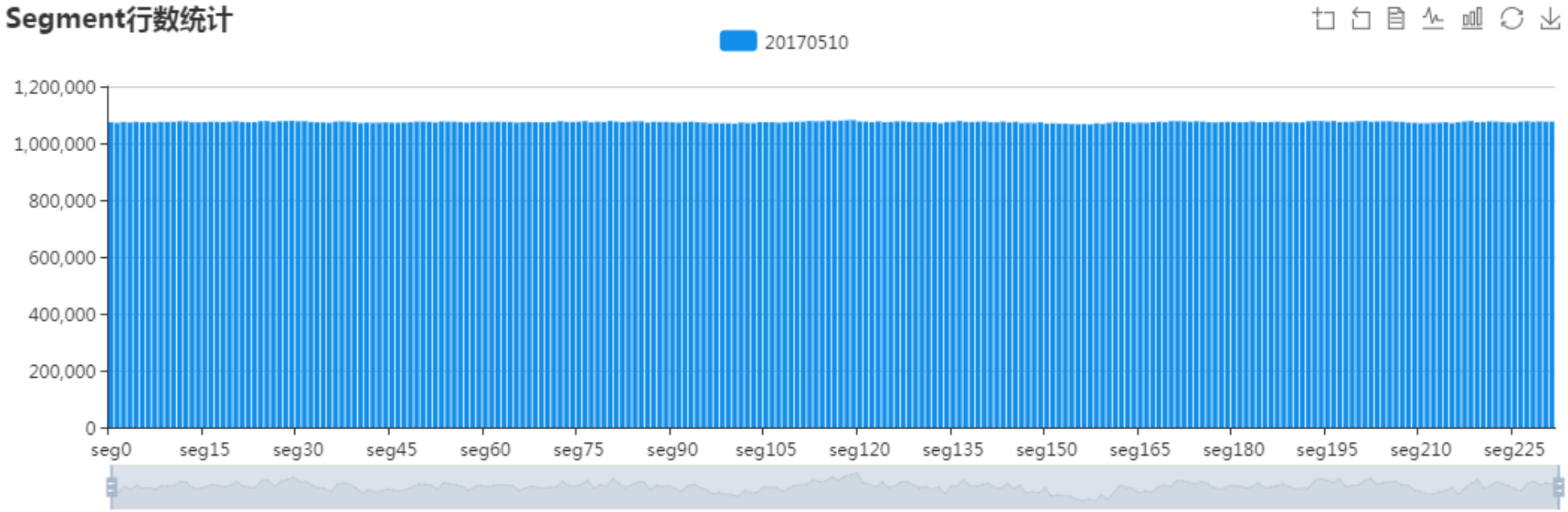


重要性:

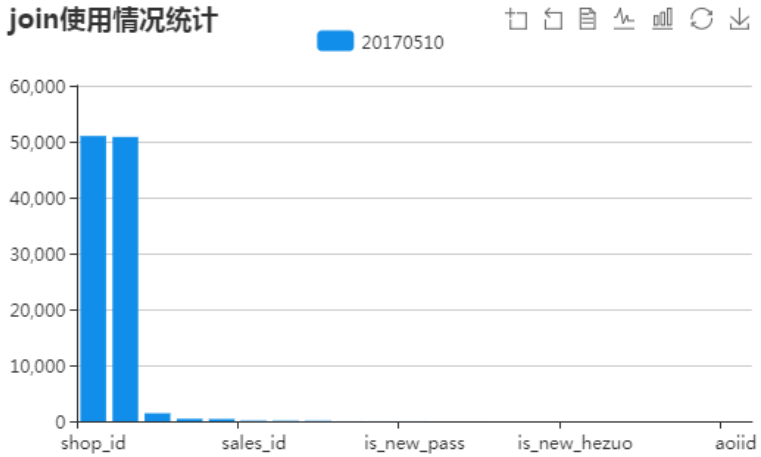
- 导致极个别节点存储资源不足，影响集群的稳定性
- sql执行效率低下
  - 最慢的节点会成为系统的瓶颈
  - 会导致不必要的广播和重分布

分布键的管理与选择

- 监控



- 自动化推荐



百度外卖研发中心

策略：

- 选择需要join的列，优先考虑高并发的列
- 结合经验与历史数据，进行统计分析，避免数据倾斜
- 避免条件字段
- 当以上条件冲突时，考虑随机分布

重要性：线上资源永远不够

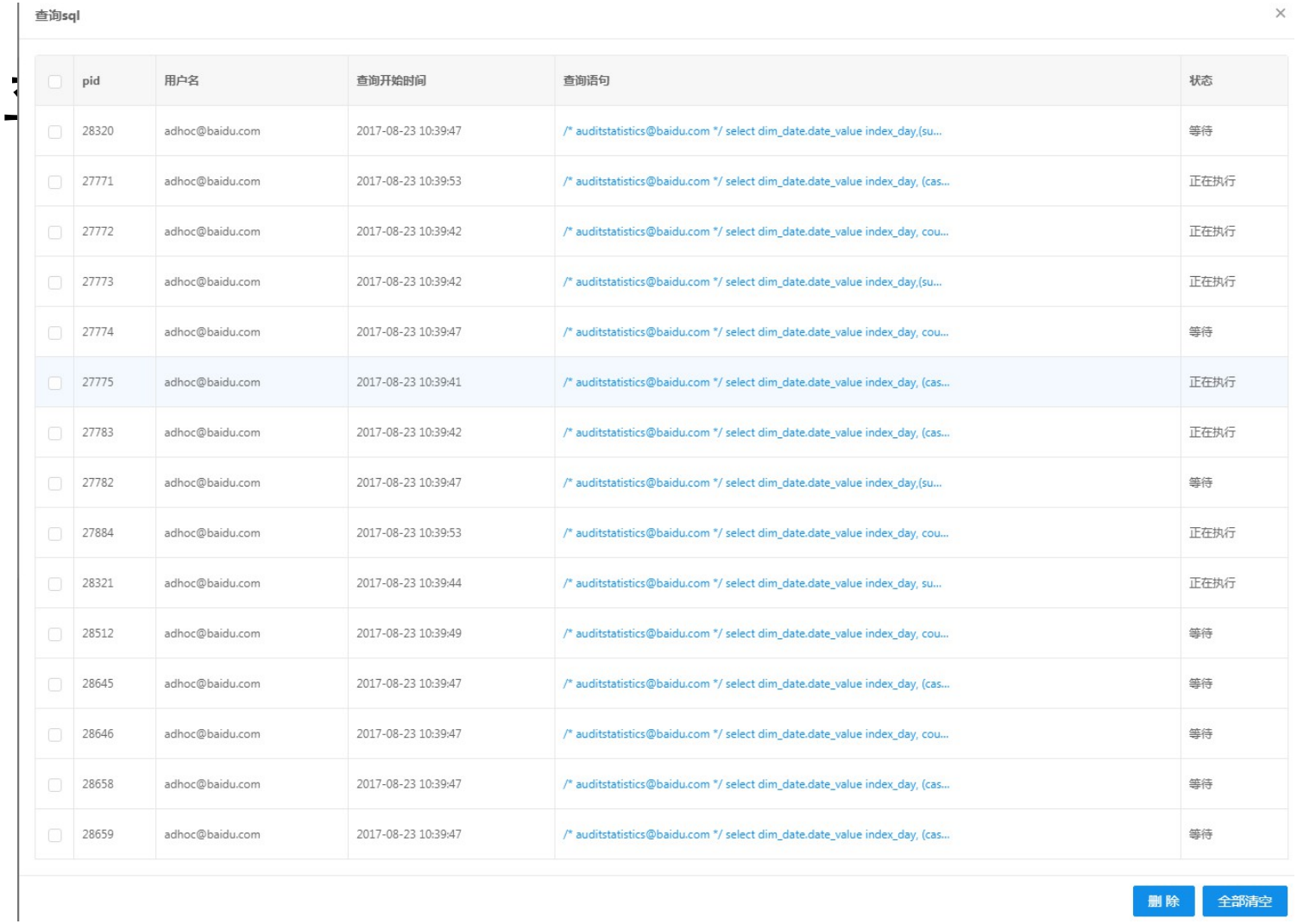
目标：

- 按照业务重要性分配角色与资源
- 通过例行动态变更，业务错峰使用资源
- 应对紧急场景，快速变更资源分配情况
- 查看与取消队列中sql查询，方便实时控制资源使用

平台化:

名称	优先级	sql数据阈值	总cost最大阈值	空闲时是否允许突破最大cost	内存限制	最小不检查cost	当前sql数目	角色组	操作
pg_default	medium	20	-1	否	1GB	0	0	gpadmin;gpmon;gpadmin_da;do@baidu.com	查询sql 操作角色组
production	medium	20	-1	否	7000MB	0	1	gpdb_optimize;gploader;doetlgpdb@baidu.com;doetlsqoop@baidu.com;appchannel	删除 修改 查询sql 操作角色组
adhoc	medium	100	1e+09	否	3000MB	0	0	renxing@baidu.com;adhoc@baidu.com	删除 修改 查询sql 操作角色组

资源队列的管理与动态变更



资源队列的管理与动态变更

Pivotal | IT大咖说 知识分享平台

平台化：

背景：

- 有一定实时场景需要支持
- 生产/消费慢查询会导致查询积压
- 分布式场景下会出现查询已经结束，但是部分节点锁未释放

锁冲突关系表:

锁模式	ACCESS SHARE	ROW SHARE	ROW EXCLUSIVE	SHARE UPDATE EXCLUSIVE	SHARE	SHARE ROW EXCLUSIVE	EXCLUSIVE	ACCESS EXCLUSIVE
ACCESS SHARE							×	×
ROW SHARE							×	×
ROW EXCLUSIVE					×	×	×	×
SHARE UPDATE EXCLUSIVE				×	×	×	×	×
SHARE			×	×	×	×	×	×
SHARE ROW EXCLUSIVE			×	×	×	×	×	×
EXCLUSIVE		×	×	×	×	×	×	×
ACCESS EXCLUSIVE	×	×	×	×	×	×	×	×

锁冲突关系表（续）：

sql语句	select,analyze	select for share	insert, copy	VACUUM (without FULL)	create index	无触发	UPDATE, select for update, DELETE	ALTER TABLE, DROP TABLE, REINDEX, CLUSTER, and VACUUM FULL，Lock命令的默认情况
select,analyze							×	×
select for share							×	×
insert, copy					×	×	×	×
VACUUM (without FULL)				×	×	×	×	×
create index			×	×	×	×	×	×
无触发			×	×	×	×	×	×
UPDATE, select for update, DELETE		×	×	×	×	×	×	×
ALTER TABLE, DROP TABLE, REINDEX, CLUSTER, and VACUUM FULL，Lock命令的默认情况	×	×	×	×	×	×	×	×

Pivotal | IT大咖说 知识分享平台

平台化:

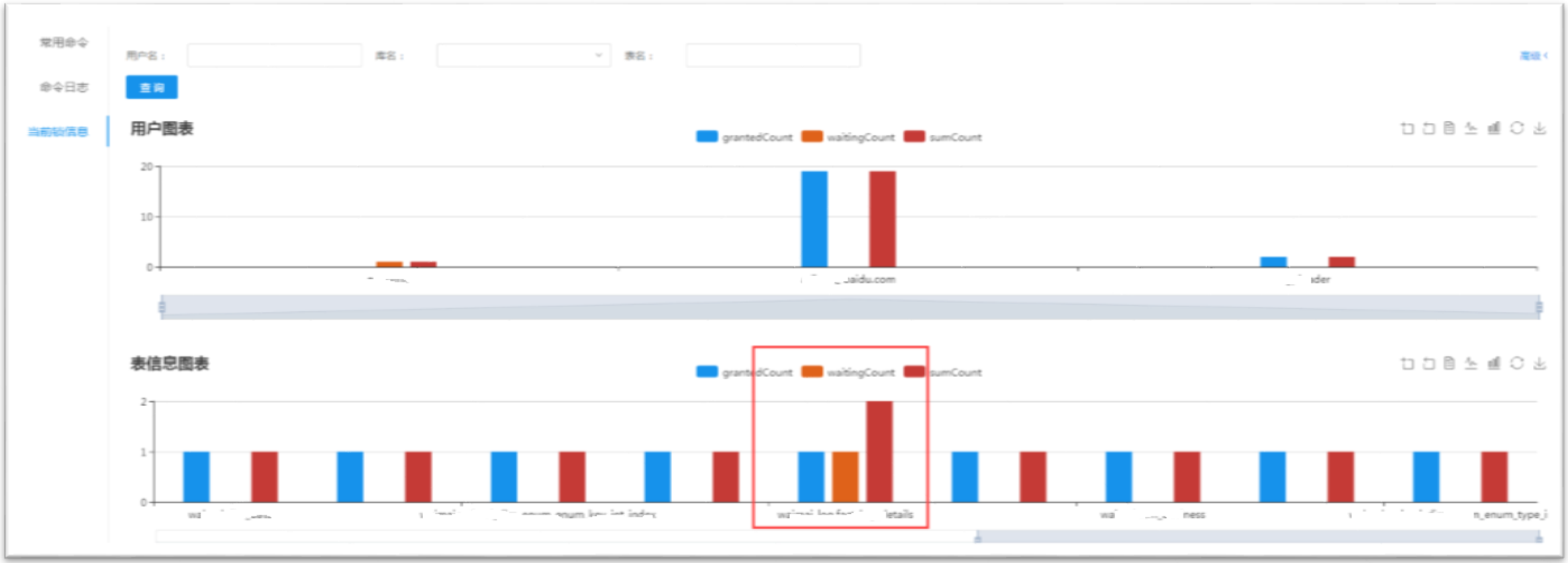
用户名： 库名： 表名： 高级

客户端IP： 输入多个以逗号分隔 资源队列： 锁模式：

锁类型： SegmentId： -1 SessionId：

Pivotal | IT大咖说 知识分享平台

平台化:



分析角度：

- sql解析，分析表/字段使用情况
- cost 高的sql分析
- 高频sql

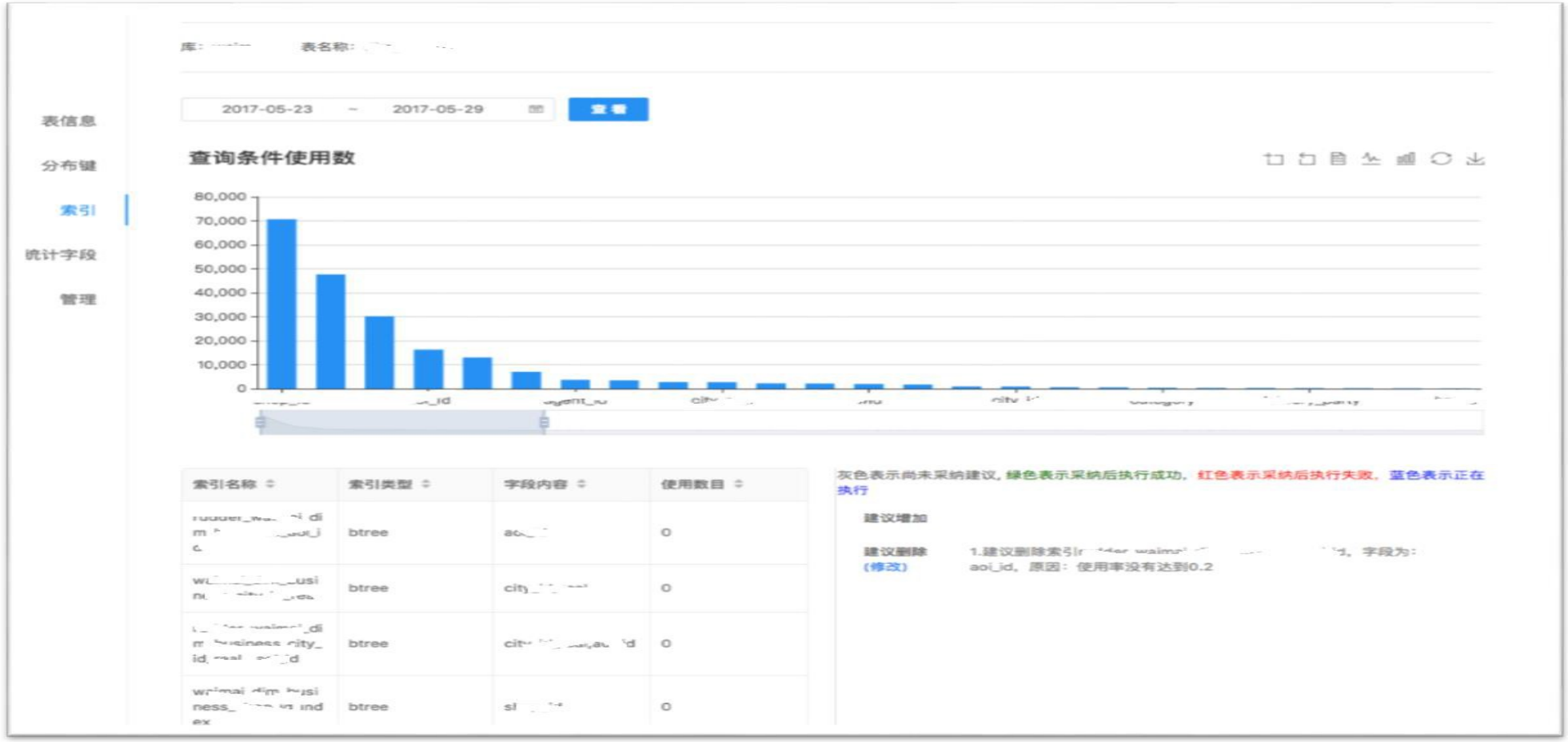
应用场景：

- 统计字段
  - 对于未使用到的字段不统计
- 索引
  - 条件筛选强的使用hash索引
  - 区分区高的使用b-tree索引
- 人工优化高频与高cost sql，并收集给出指导意见

围绕sql分析的优化

Pivotal | IT大咖说 知识分享平台

平台化

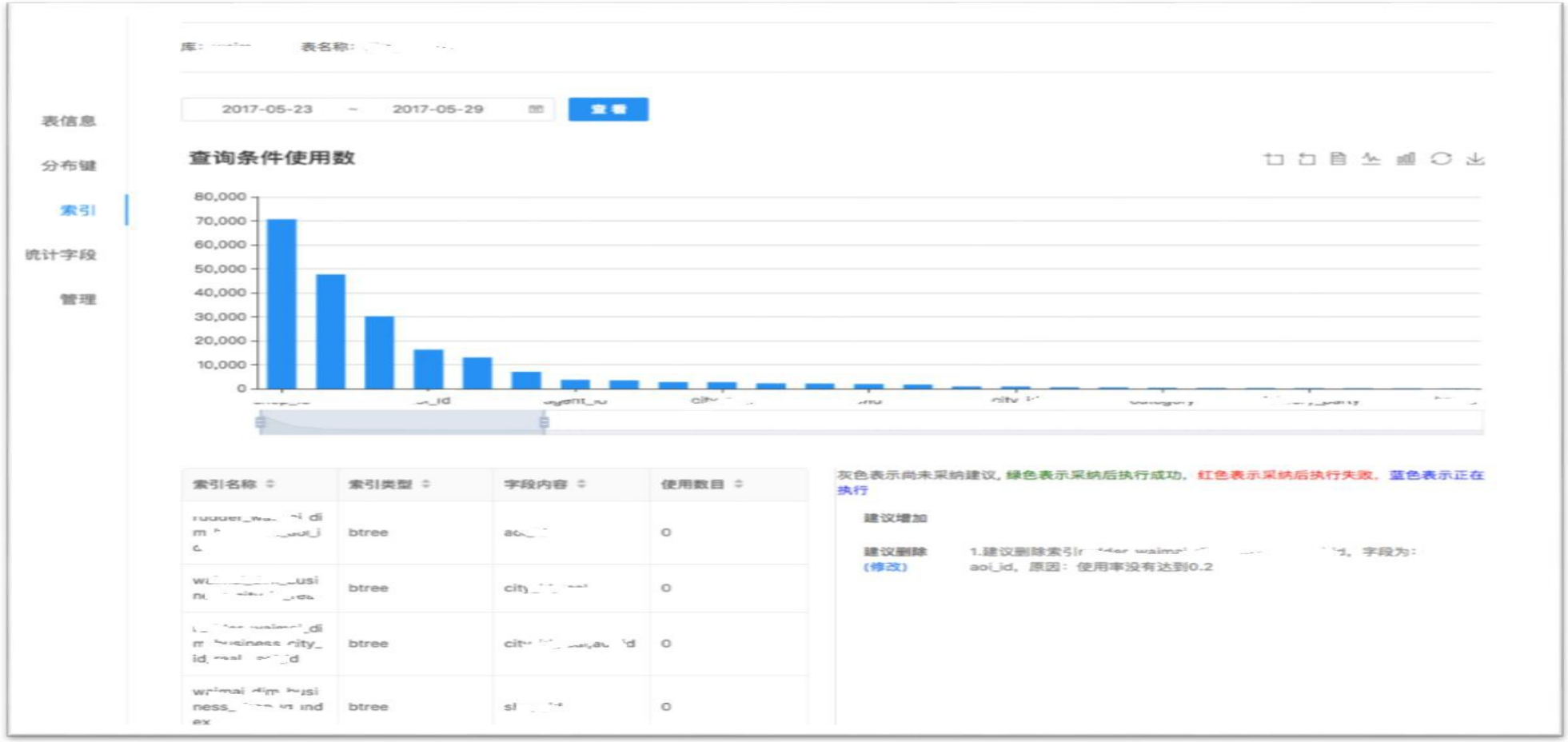


百度外卖研发中心

围绕sql分析的优化

Pivotal | IT大咖说 知识分享平台

平台化



百度外卖研发中心

围绕sql分析的优化

Pivotal | IT大咖说 知识分享平台

平台化

聚合方式： 按平均值 时间： 2017-05-23 00:00:00 ~ 2017-05-28 00:00:00 查询

adhocId	花费时间	默认Schema	查询执行时间	结果数目	用户名	查询语句	sqlInfold	操作
41683335	10h	[redacted]	2017-05-27 08:34:02	25	[redacted]@baidu.com	select dim_date.d...	1697573	查看
41683283	10h	[redacted]	2017-05-27 08:33:17	8	[redacted]@baidu.com	select dim_date.d...	1697533	查看
41683284	10h	[redacted]	2017-05-27 08:33:17	8	[redacted]baidu.com	select dim_date.d...	1697534	查看
41683293	10h	[redacted]	2017-05-27 08:33:20	2	[redacted]@baidu.com	select dim_date.d...	1697543	查看
41683286	10h	[redacted]	2017-05-27 08:33:17	8	[redacted]tics@baidu.com	select dim_date.d...	1697536	查看
41683337	10h	[redacted]	2017-05-27 08:34:03	8	[redacted]@baidu.com	select dim_date.d...	1697575	查看
41683327	10h	[redacted]	2017-05-27 08:33:56	0	[redacted]@baidu.com	select dim_date.d...	1697567	查看
41683277	10h	[redacted]	2017-05-27 08:33:16	1	[redacted]@baidu.com	select dim_date.d...	1697528	查看

百度外卖研发中心

天级别

- 修改分区表的存储方式（冷热数据）
- 例行vacuum，清理膨胀表
- analyze个别大分区表

分钟/小时级别

- 删除idle session
- 删除超时查询
- 拉起down掉的节点
- 分布不均匀的节点，重分布
- 监控节点残留锁并删除

- 承担更多的业务与场景
- sql解析工具更加专业化
- 解决内存溢出的问题
- 收集gp执行日志进行详细分析，监控资源使用情况

Pivotal | IT大咖说 知识分享平台

谢谢

2017.08

百度外卖研发中心