

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

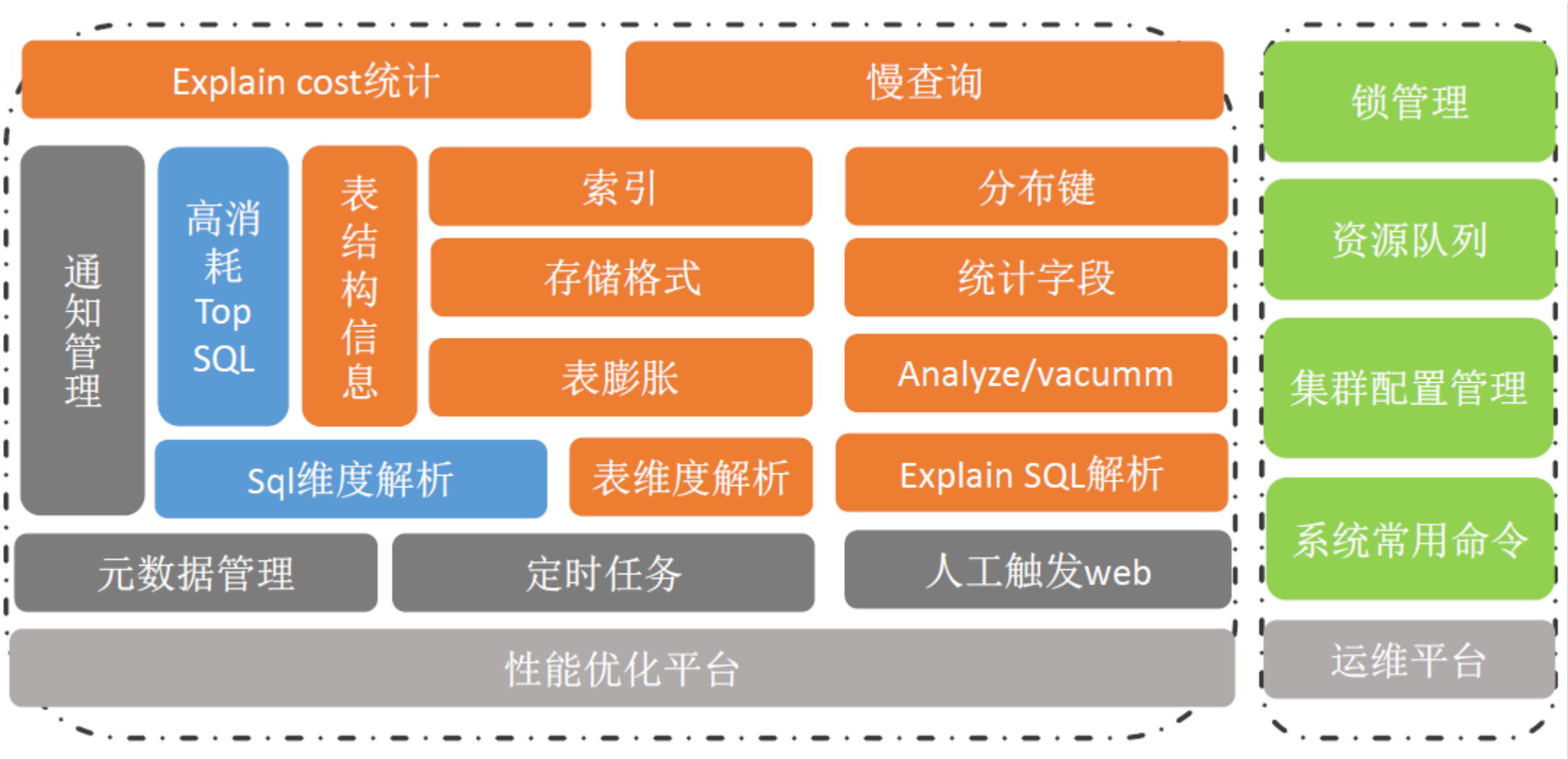
- 为什么平台化
- 分布键管理与选择
- 资源队列管理与动态变更
- 锁管理与深入分析
- 围绕sql分析的优化
- 一些例行运维操作
- 未来与挑战

- 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

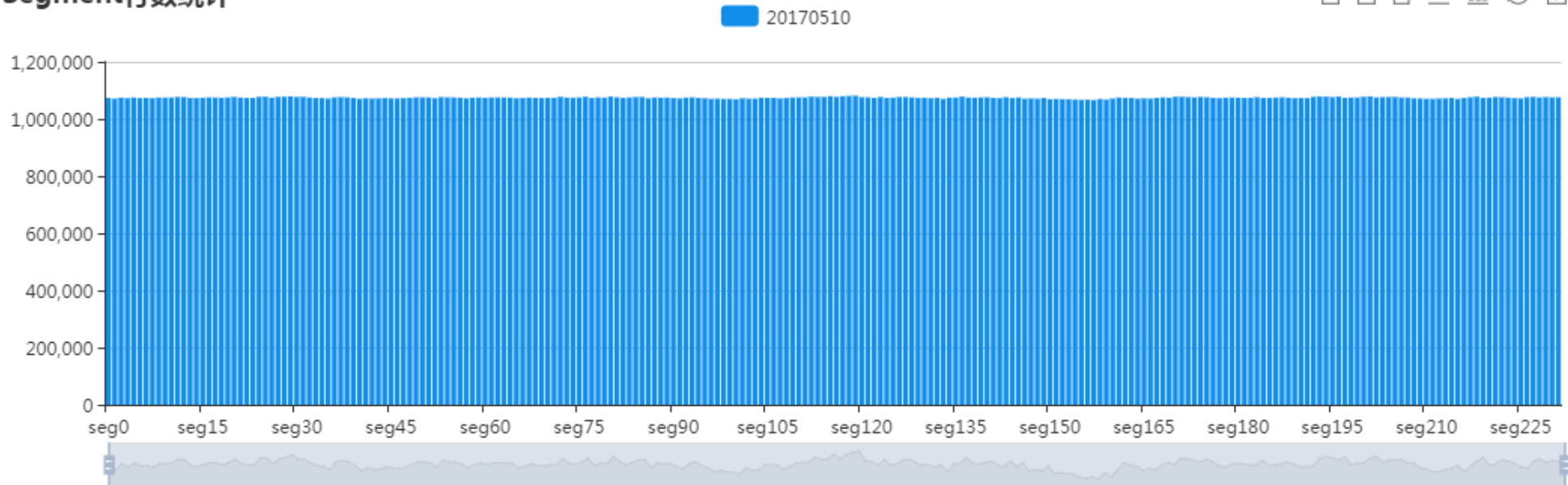


重要性：

- 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

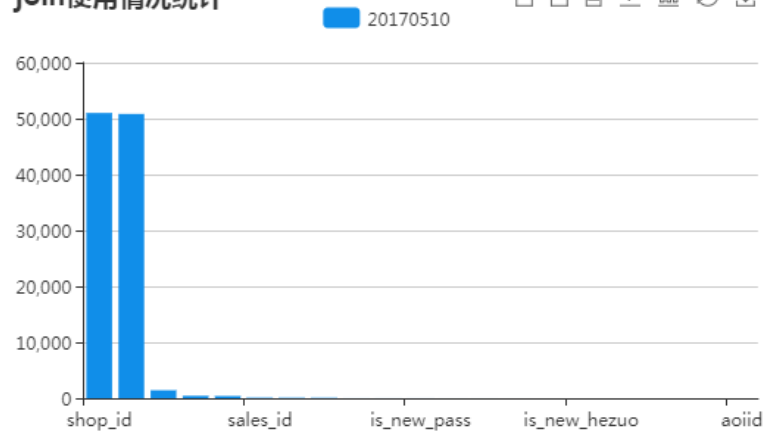
● 监控

Segment行数统计



● 自动化推荐

join使用情况统计



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	gpgadmin;gpmon;gpgadmin_da;do@bai du.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;ap pchannel	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.co m	删除 修改 查询sql 操作角色组

查询sql

☐	pid	用户名	查询开始时间	查询语句	状态
☐	28320	adhoc@baidu.com	2017-08-23 10:39:47	/* auditstatistics@baidu.com */ select dim_date.date_value index_day(su...	等待
☐	27771	adhoc@baidu.com	2017-08-23 10:39:53	/* auditstatistics@baidu.com */ select dim_date.date_value index_day, (cas...	正在执行
☐	27772	adhoc@baidu.com	2017-08-23 10:39:42	/* auditstatistics@baidu.com */ select dim_date.date_value index_day, cou...	正在执行
☐	27773	adhoc@baidu.com	2017-08-23 10:39:42	/* auditstatistics@baidu.com */ select dim_date.date_value index_day(su...	正在执行
☐	27774	adhoc@baidu.com	2017-08-23 10:39:47	/* auditstatistics@baidu.com */ select dim_date.date_value index_day, cou...	等待
☐	27775	adhoc@baidu.com	2017-08-23 10:39:41	/* auditstatistics@baidu.com */ select dim_date.date_value index_day, (cas...	正在执行
☐	27783	adhoc@baidu.com	2017-08-23 10:39:42	/* auditstatistics@baidu.com */ select dim_date.date_value index_day, (cas...	正在执行
☐	27782	adhoc@baidu.com	2017-08-23 10:39:47	/* auditstatistics@baidu.com */ select dim_date.date_value index_day(su...	等待
☐	27884	adhoc@baidu.com	2017-08-23 10:39:53	/* auditstatistics@baidu.com */ select dim_date.date_value index_day, cou...	正在执行
☐	28321	adhoc@baidu.com	2017-08-23 10:39:44	/* auditstatistics@baidu.com */ select dim_date.date_value index_day, su...	正在执行
☐	28512	adhoc@baidu.com	2017-08-23 10:39:49	/* auditstatistics@baidu.com */ select dim_date.date_value index_day, cou...	等待
☐	28645	adhoc@baidu.com	2017-08-23 10:39:47	/* auditstatistics@baidu.com */ select dim_date.date_value index_day, (cas...	等待
☐	28646	adhoc@baidu.com	2017-08-23 10:39:47	/* auditstatistics@baidu.com */ select dim_date.date_value index_day, cou...	等待
☐	28658	adhoc@baidu.com	2017-08-23 10:39:47	/* auditstatistics@baidu.com */ select dim_date.date_value index_day, (cas...	等待
☐	28659	adhoc@baidu.com	2017-08-23 10:39:47	/* auditstatistics@baidu.com */ select dim_date.date_value index_day, (cas...	等待

删除

全部清空

平台化：

当前列表

动态计划

选择创建资源队列集合模板:

集合名称	时间类型	日期范围	开始时间	结束时间	默认	操作
default	-	-	-	-	是	查看集合详细信息
生产队列	天类型	-	00:04	09:00	否	删除资源队列集合 查看集合详细信息 修改资源队列计划集合

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							x	x
ROW SHARE							x	x
ROW EXCLUSIVE					x	x	x	x
SHARE UPDATE EXCLUSIVE				x	x	x	x	x
SHARE			x	x	x	x	x	x
SHARE ROW EXCLUSIVE			x	x	x	x	x	x
EXCLUSIVE		x	x	x	x	x	x	x
ACCESS EXCLUSIVE	x	x	x	x	x	x	x	x

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							x	x
select for share							x	x
insert, copy					x	x	x	x
VACUUM (without FULL)				x	x	x	x	x
create index			x	x	x	x	x	x
无触发			x	x	x	x	x	x
UPDATE, select for update, DELETE		x	x	x	x	x	x	x
ALTER TABLE, DROP TABL E, REINDEX, CLUSTER, and VACUUM FULL, Lock命令的 默认情况	x	x	x	x	x	x	x	x

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：

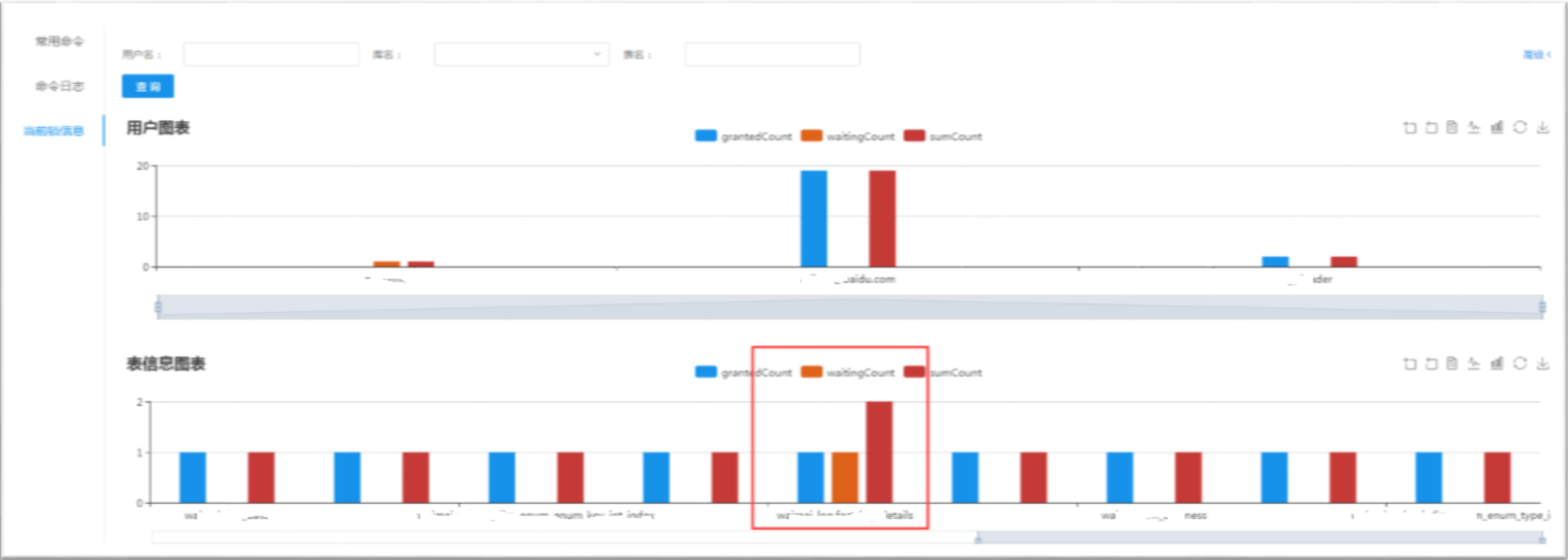
详情信息

库名	表名	锁模式	进程id	用户名	始时间	segmentid	查看	操作
		AccessExclusiveLock	6409		06:03	-1	被本锁阻塞的锁 导致本锁阻塞的锁	删除SessionID

取消

确定

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

[illegible]

[illegible]

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		2017-05-27 08:34:02	25		select dim_date.d..	1697573	查看
41683283	10h		2017-05-27 08:33:17	8		select dim_date.d..	1697533	查看
41683284	10h		2017-05-27 08:33:17	8		select dim_date.d..	1697534	查看
41683293	10h		2017-05-27 08:33:20	2		select dim_date.d..	1697543	查看
41683286	10h		2017-05-27 08:33:17	8		select dim_date.d..	1697536	查看
41683337	10h		2017-05-27 08:34:03	8		select dim_date.d..	1697575	查看
41683327	10h		2017-05-27 08:33:56	0		select dim_date.d..	1697567	查看
41683277	10h		2017-05-27 08:33:16	1		select dim_date.d..	1697528	查看

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

谢谢



2017.08

百度外卖研发中心