



数据技术嘉年华

Data Technology Carnival

云·数据·智能 - 数聚价值智胜未来

SQL优化利器—SQLADVISOR

美团点评技术专家 — 龙雪刚



第七屆



数据技术嘉年华

Data Technology Carnival



大纲

- 1. 个人介绍
- 2. SQLAdvisor介绍
- 3. SQLAdvisor使用
- 4. 未来计划



第七屆



数据技术嘉年华
Data Technology Carnival



个人简介

- 餐饮酒旅DBA团队Leader（技术专家）
- 负责外卖、酒旅、打车、电影、团购业务MySQL的技术保障服务
- 曾任职于京东等，7年专职DBA经验，Devops实践者



第七屆



数据技术嘉年华
Data Technology Carnival



招聘：DBA+SRE

- 为**gh-ost**贡献**issue+patch**
- 为**mha**贡献**issue+patch** 开源的
- **dbproxy**、**sqladvisor**、**zebra**、**star**数超过**3000+**
- **Cellar**计划开源(基础架构部开发)
- 负责外卖、配送、金融、酒旅、打车、零售、猫眼各业务线
- 邮箱：longxeugang@meituan.com (电话/微信13661059141)



第七屆



数据技术嘉年华
Data Technology Carnival



大纲

- 1. 个人介绍
- 2. SQLAdvisor介绍
- 3. SQLAdvisor使用
- 4. 未来计划



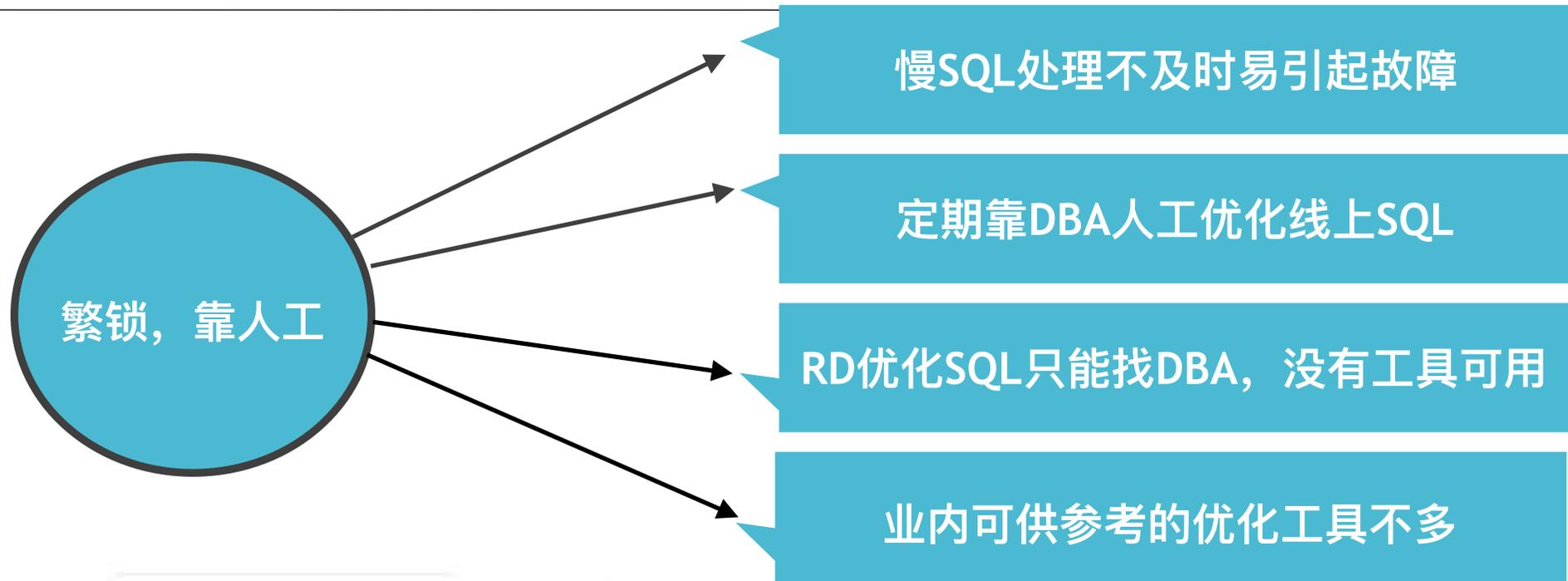
第七屆



数据技术嘉年华
Data Technology Carnival



背景



第七屆

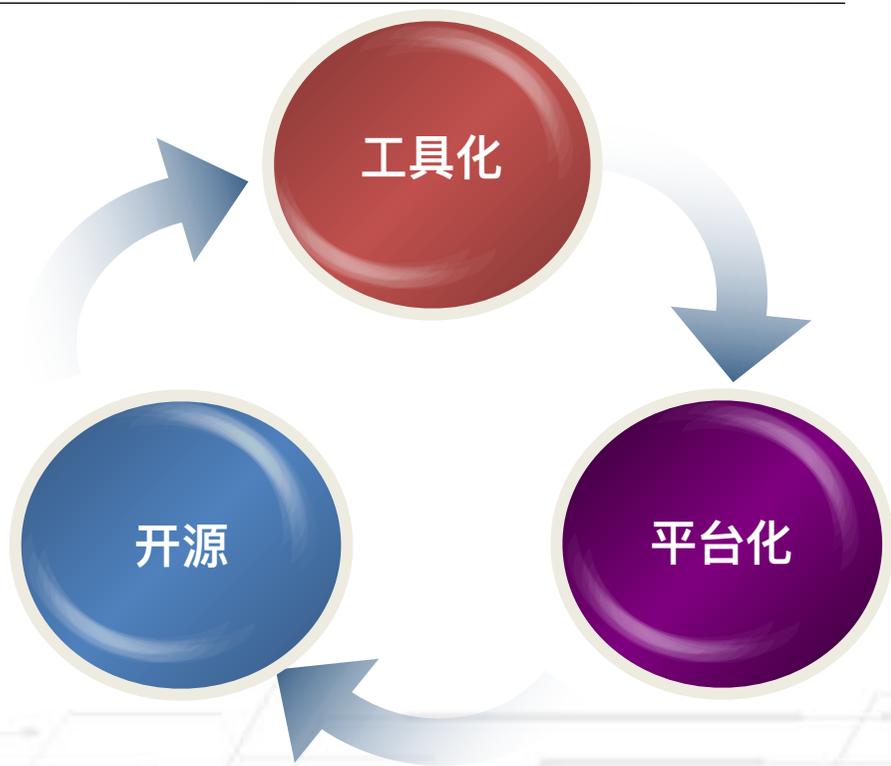


数据技术嘉年华
Data Technology Carnival

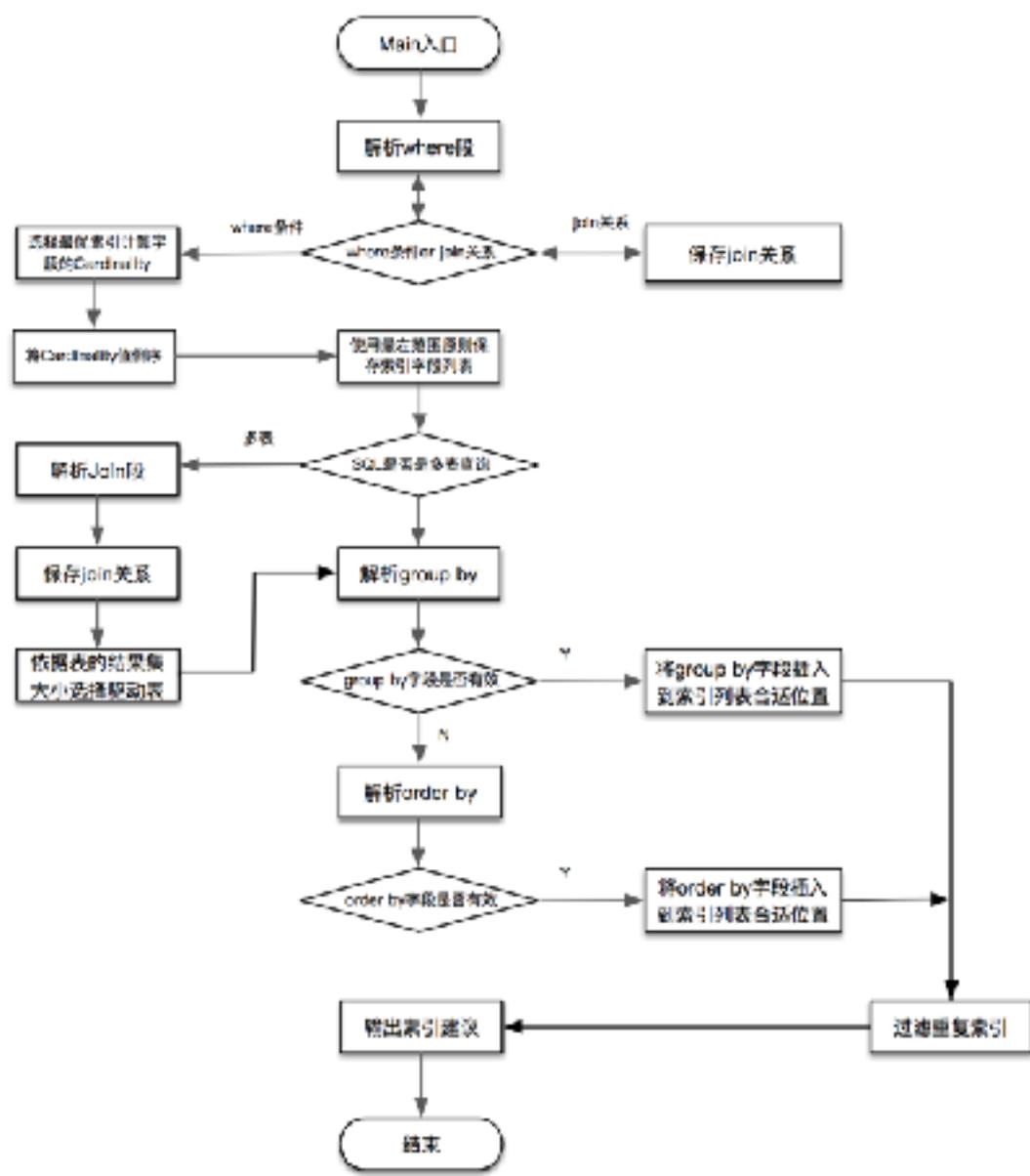


目标：SQL自助优化、解放DBA

工具化	<ul style="list-style-type: none">设计整个架构，技术难点逐个突破优化代码、解决使用中的问题
平台化	<ul style="list-style-type: none">与DBA平台打通、RD自助优化搜集线上慢SQL，自动给出建议
开源	<ul style="list-style-type: none">工具开源，借助开源力量继续优化扩大公司品牌，吸引人才



整体架构



Where处理

1. 只处理 AND 连接符，忽略OR。
2. where 条件中可以存在 Join 条件，需要进行区分。
3. 操作符like，如果不是前缀匹配则丢弃这个条件。
4. 区分度小于30则丢弃，按照高低进行倒序排。



第七屆



数据技术嘉年华
Data Technology Carnival

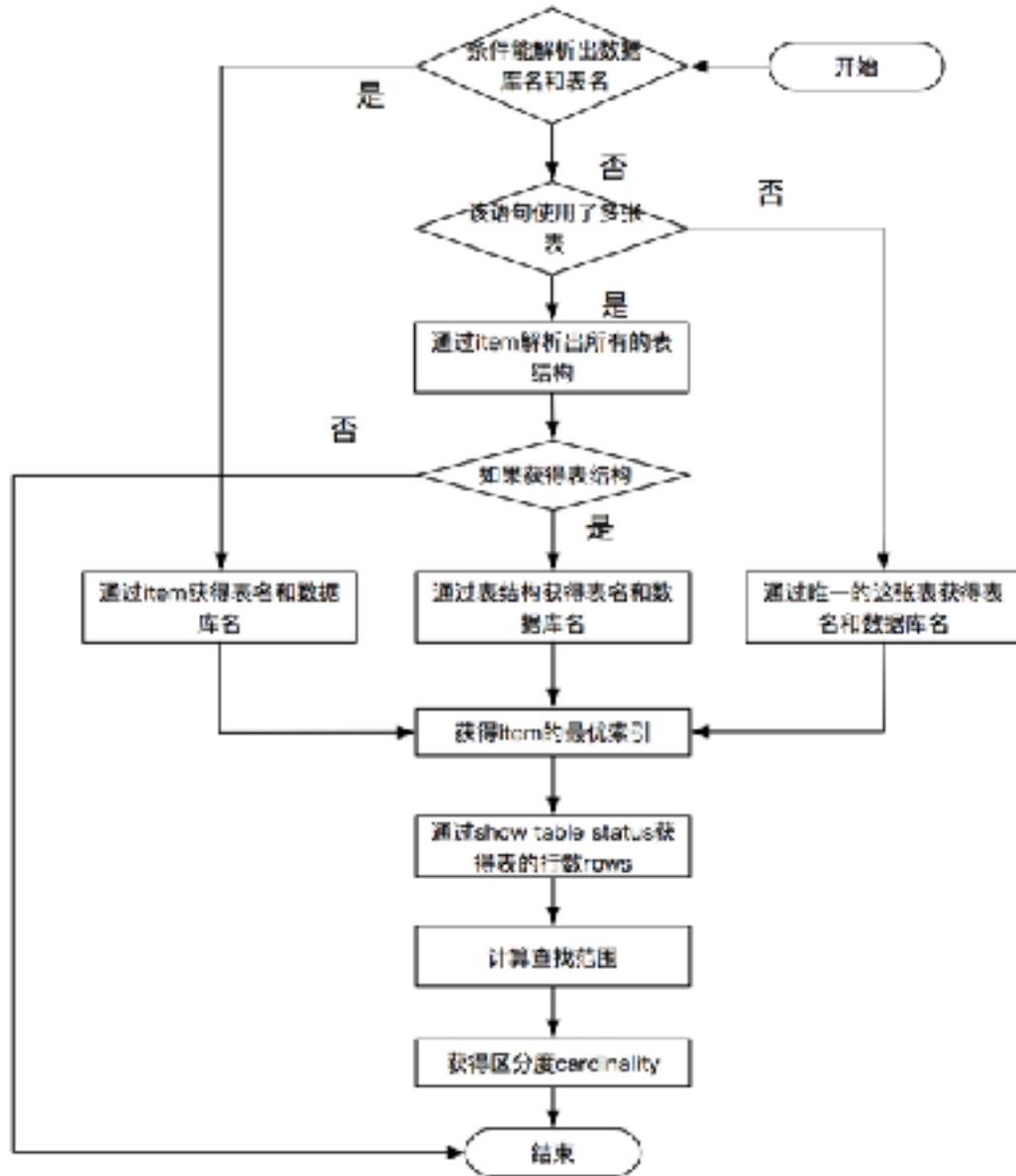


计算区分度

1. “show table status like” 获得表的总行数 table_count。
2. 选择表中已存在的区分度最高的索引 best_index, 同时Primary key > Unique key > 一般索引。
3. 计算数据采样的起始值offset与采样范围rand_rows:
 - $offset = (table_count / 2) > 10W ? 10W : (table_count / 2)$
 - $rand_rows = (table_count / 2) > 1W ? 1W : (table_count / 2)$
 - 使用select count(1) from (select field from table force index(best_index) order by cl.. desc limit rand_rows) where field_print 得到满足条件的rows。
 - $cardinality = rows == 0 ? rand_rows : rand_rows / rows;$
4. 计算完成选择度后, 会根据选择度大小, 将该条件添加到该表中的备选索引中。



计算区分度

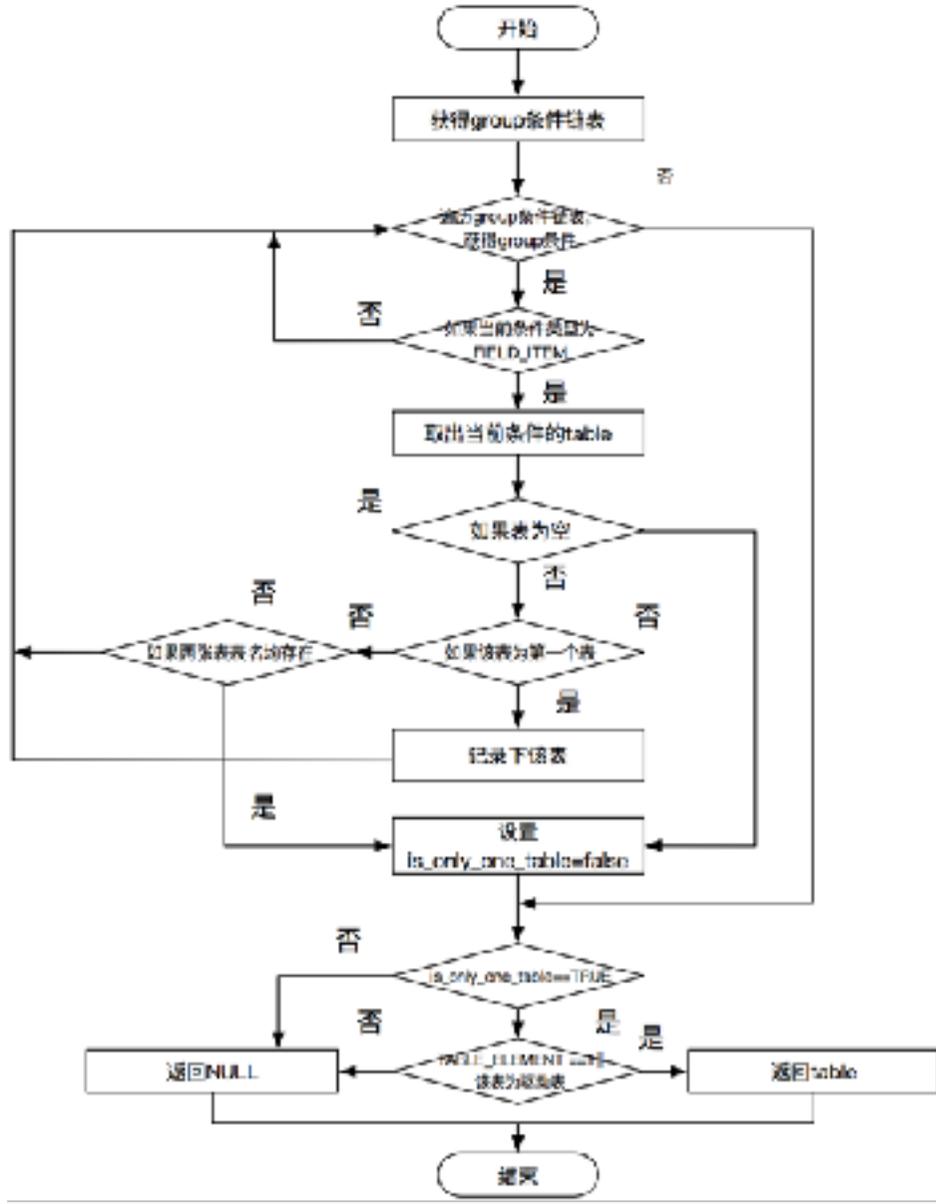


Group By & Order By

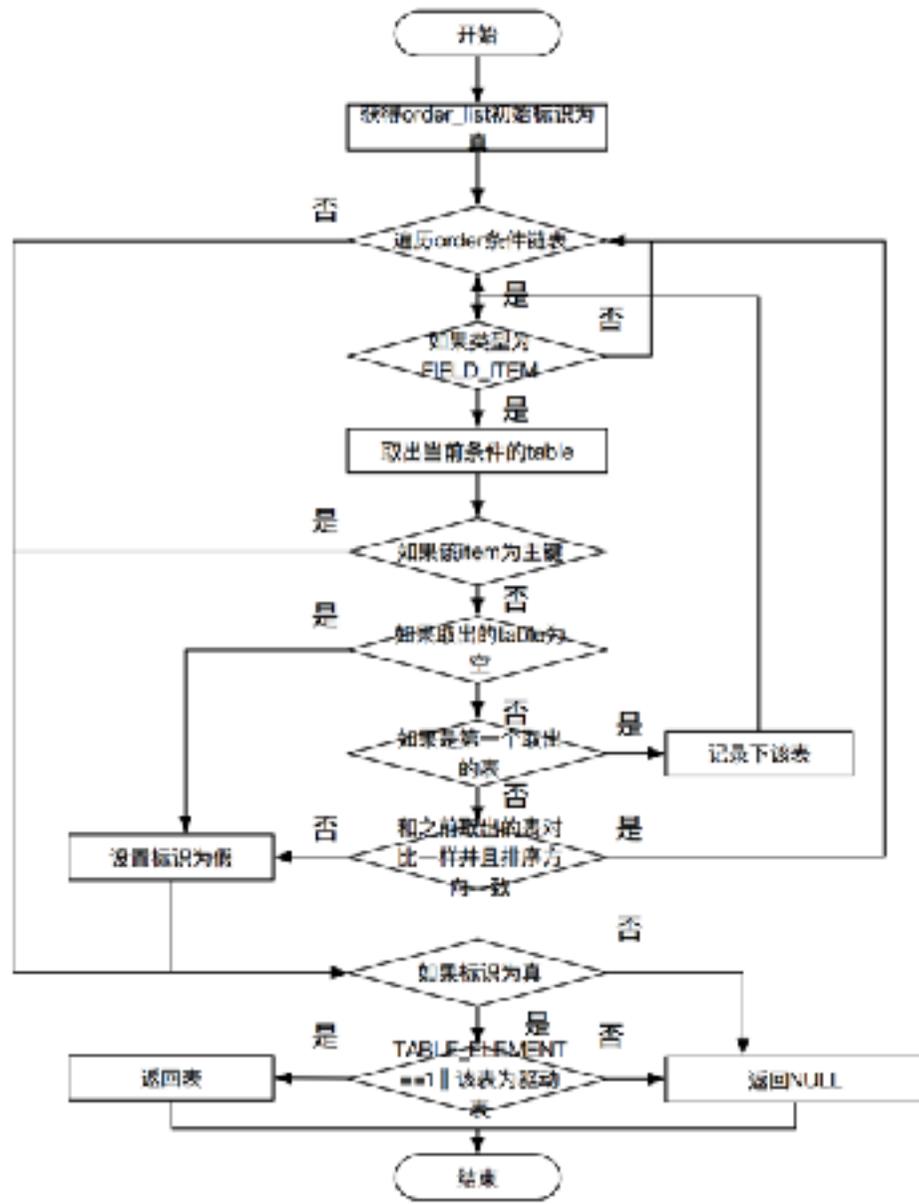
1. Group 字段与 Order 字段能否用上索引，需要满足如下条件：
 - 涉及的字段必须来自于同一张表，并且这张表必须是确定下来的驱动表。
 - Group by 优于 Order by，两者只能同时存在一个。
 - Order by 字段的排序方向必须完全一致，否则丢弃整个 Order by 字段列。
 - 当 Order by 条件中包含主键时，如果主键字段为 Order by。 字段列末尾，忽略该主键，否则丢弃整个 Order by 字段列。
2. 整个索引列排序优先级：等值>(group by | order by)> 非等值。



GroupBy



Order By

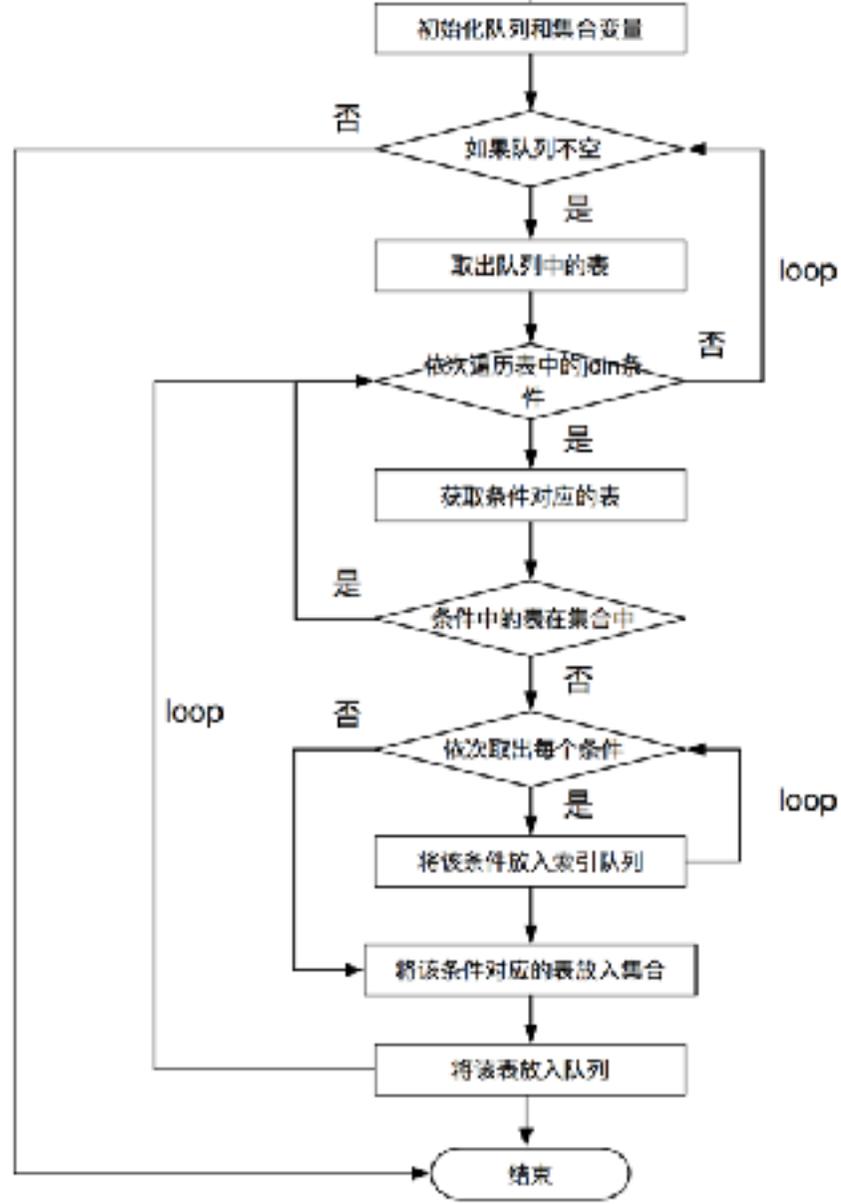


选择驱动表

1. 经过前期 where 解析、Join 解析，已经将关联关系存储起来，并且按照一定逻辑将候选驱动表确定下来。
2. 在候选驱动表中，按照每一张表的候选索引字段中第一个字段进行计算表中结果集大小。
3. 使用 `explain select * from table where field` 来计算表中结果集。
4. 结果集最小的被确为驱动表。



选择驱动表



结果输出

1. 过滤重复索引
2. 该步骤涉及的函数是: `print_index()`



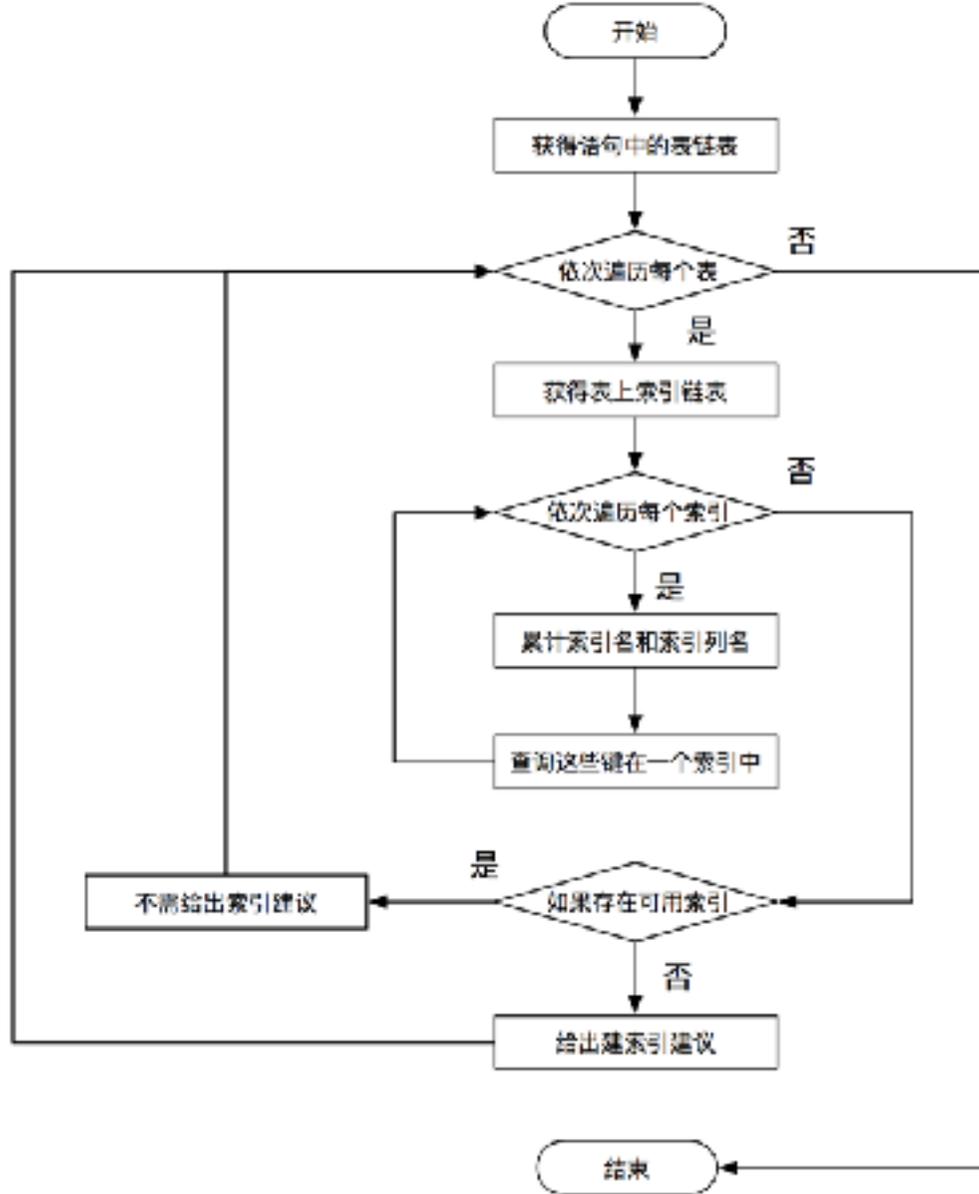
第七屆



数据技术嘉年华
Data Technology Carnival



结果输出



大纲

- 1. 个人介绍
- 2. SQLAdvisor介绍
- 3. SQLAdvisor使用
- 4. 未来计划



第七屆



数据技术嘉年华
Data Technology Carnival



SQL: `SELECT DISTINCT * FROM skinfo_item where airCo='KN' AND dstap='NNY' AND dstterm!='-' AND iscodeshare=0`

输入SQL

经过SQLAdvisor优化

输出SQL建议

SQL建议: SQL中使用了SELECT *。尽量指定列名,而不是用*号

索引建议: 2017-03-22 15:42:41 25300 [Note] Create_Index_SQL: alter table skinfo_item add index idx_airCo_dstap(airCo,dstap)

分析日志: [打开详情](#)

输出索引建议



第七屆



数据技术嘉年华
Data Technology Carnival



大纲

- 1. 个人介绍
- 2. SQLAdvisor介绍
- 3. SQLAdvisor使用
- 4. 未来计划



第七屆



数据技术嘉年华
Data Technology Carnival



未来计划

- 优化词法解析：提高SQL解析准确率
- 区分度计算优化：参考5.7 COST算法
- 年底更新一版
- 解决BUG...



第七屆



数据技术嘉年华
Data Technology Carnival





THANKS



招聘：DBA+SRE

- 为**gh-ost**贡献**issue+patch**
- 为**mha**贡献**issue+patch** 开源的
- **dbproxy**、**sqladvisor**、**zebra**、**star**数超过**3000+**
- **Cellar**计划开源(基础架构部开发)
- 负责外卖、配送、金融、酒旅、打车、零售、猫眼各业务线
- 邮箱：longxeugang@meituan.com (电话/微信13661059141)



第七屆



数据技术嘉年华
Data Technology Carnival

