MySQL 8.0反连接(antijoin) 原创

mb5ff981d806017 2021-03-01 19:21:37

文章标签 java 文章分类 Java 后端开发 阅读数 512

引言

在MySQL8.0.16版本之前，对于IN和EXISTS的处理，优化器可以将IN由子查询方式优化为semi join,但EXITS只能采用子查询的方式，所以在执行计划中看到的是DEPENDENT SUBQUERY。在8.0.16版本中，对EXISTS进行了优化，使其可以像IN一样支持转换为semi join:

> Beginning with MySQL 8.0.16, the semijoin optimizations used with IN subqueries can now be applied to EXISTS subqueries as well
> https://dev.mysql.com/doc/refman/8.0/en/mysql-nutshell.htm

在MySQL8.0.17版本中，对NOT IN,NOT EXISTS进行了优化，将其转变为anti join:

> In short, any negation of a subquery of the form IN (SELECT ... FROM ...) or EXISTS (SELECT ... FROM ...) is transformed into an antijoin
> https://dev.mysql.com/doc/refman/8.0/en/semijoins.html

环境准备：

1. 192.168.56.101 MySQL8.0.20192.168.56.102 MySQL5.7.19192.168.56.103 SYSBENCHdatabase:sysbenchtables:sbtest1,sbtest2

表结构如下均一致：

```
mysql> desc sbtest1;
+-------+----------+------+-----+---------+----------------+
| Field | Type     | Null | Key | Default | Extra          |
+-------+----------+------+-----+---------+----------------+
| id    | int(11)  | NO   | PRI | NULL    | auto_increment |
| k     | int(11)  | NO   | MUL | 0       |                |
| c     | char(120)| NO   |     |         |                |
| pad   | char(60) | NO   |     |         |                |
+-------+----------+------+-----+---------+----------------+
4 rows in set (0.00 sec)
```

数据量：

```
mysql> select count(*) from sbtest1;
+----------+
| count(*) |
+----------+
|   999990 |
+----------+
1 row in set (0.10 sec)

mysql> select count(*) from sbtest2;
+----------+
| count(*) |
+----------+
|  1000000 |
+----------+
1 row in set (0.10 sec)
```

操作明细：

5.7.19中的IN操作：

1. SQL>explain select * from sbtest1 a where a.id in (select b.id from sbtest2 b);

可见IN操作转换为了semi join

5.7.19中的EXISTS操作：

1. SQL>explain select * from sbtest1 a where exists (select 1 from sbtest2 b where a.id=b.id);

可见EXISTS操作在5.7.19中没有做优化，依然为子查询方式

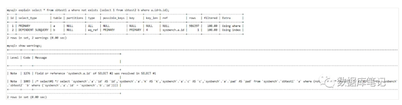
5.7.19中的NOT IN操作：

1. SQL>explain select * from sbtest1 a where a.id not in (select b.id from sbtest2 b);

可见5.7.19中NOT IN操作没有做优化处理

5.7.19中的NOT EXISTS操作:

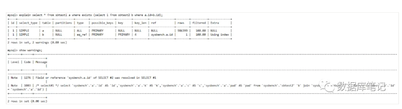
1. SQL>explain select * from sbtest1 a where not exists (select 1 from sbtest2 b where a.id=b.id);



可见5.7.19中NOT EXISTS操作没有做优化处理

8.0.20中的IN操作：

1. SQL>explain select * from sbtest1 a where a.id in (select b.id from sbtest2 b);



与5.7.19一致

8.0.20中的EXISTS操作：

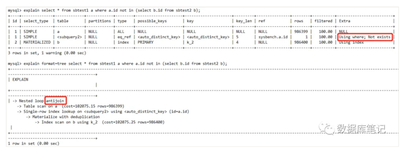
1. SQL>explain select * from sbtest1 a where exists (select 1 from sbtest2 b where a.id=b.id);

可见，8.0.16版本后，EXISTS操作与IN操作优化一致

8.0.20中的NOT IN操作：

1. SQL>explain format=tree select * from sbtest1 a where a.id not in (select b.id from sbtest2 b);

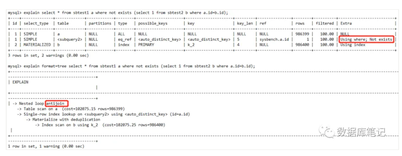
PS:使用format=tree可以查看执行计划的详细预估成本信息，这边为了看清antijoin，如果需要查看实际执行花费，可以用analyze



可见，8.0.17版本后，NOT IN采用antijoin方式进行优化

8.0.20中的NOT EXISTS操作：

1. SQL>explain format=tree select * from sbtest1 a where not exists (select 1 from sbtest2 b where a.id=b.id);



可见，8.0.17版本后，NOT EXISTS同样采用antijoin方式进行优化

结果分析：

我们知道子查询方式，需要遍历表a的记录，每一条记录都需要检索一遍表b的子查询，当数据量增大的情况下，效率会越来越差。所以优化器会将in,exists转换为semi join,将not in,not exists转换为antijoin。但我们可以看到，antijoin在执行计划中表现形式之一为MATERIALIZED（anti join的策略选择基于成本考虑有多种，具体可见https://dev.mysql.com/doc/refman/8.0/en/semijoins.html），也即是会对子查询创建一个物化临时表，会消耗一定的

mb5ff981d806017
150 原创 117.6万 人气 8 粉丝 4030 评论
0 翻译 2187 转载 0 关注 13 收藏
+ 关注 私信

近期文章
1.【小沐学前端】Node.js实现基于Proto…
2.GenericServlet 和 HttpServlet
3.Java常用类（重要 · 下篇）
4.无涯教程-Erlang - 性能
5.Linux配置Java环境变量（详细步骤总结

赞 收藏 评论 分享 举报

上一篇：MySQL8.0 clone plugin

下一篇：MySQL中存储过程，函数，触发器须知

提问和评论都可以，用心的回复会被更多人看到 评论

相关文章

MySQL 8.0反连接(antijoin)

引言 在MySQL8.0.16版本之前，对于IN和EXISTS的处理，优化器可以将IN由子查询方式优化为semi join,但EXITS只能采用…

java

java sql 反斜杠处理

Java SQL反斜杠处理## 引言在Java开发中，我们经常会遇到需要处理SQL语句的情况。其中一个常见的问题是如何处理反斜…

SQL 反斜杠 Java

SQL反模式学习笔记21 SQL注入

目标：编写SQL动态查询，防止SQL注入 通常所说的“SQL动态查询”是指将程序中的变量和基本SQL语句拼接成一个完整的查询…

SQL设计模式

SQL连接

在sql server中，我们经常能用到连接，今天总结一下连接的基础知识。连接的分类

自连接 数据 内连接

一个反直觉的sql

引子在《容易引起雪崩的两个处理》里，我提到一个慢查询的问题。本文先从整洁架构的角度讲讲慢查询sql完成的功能以及设计…

java 数据库 mysql python 索引

SQL反模式学习笔记1 开篇

什么是“反模式” 反模式是一种试图解决问题的方法，但通常会同时引发别的问题。 反模式分类 （1）逻辑数据库设计反模式 在开…

SQL设计模式

SQL反模式学习笔记15 分组

目标：查询得到每组的max（或者min等其他聚合函数）值，并且得到这个行的其他字段 反模式：引用非分组列 单值规则：跟在…

SQL设计模式

mySQL反连接 mysql反模式

反范式是试图通过增加冗余数据或通过分组数据来优化数据库读取性能的过程。在某些情况下，反范式是解决数据库性能和可伸…

mySQL反连接 反范式 join 数据 数据库

sql反模式分析1

第二章：乱穿马路 2.1 目标：存储多值属性 2.2 反模式：格式化的逗号分隔列表 模糊匹配无法使用索引，影响性能；多表关联麻…

子类 外键 数据 解决方案 主键

NULL对反连接的影响

测试准备： 如果T1表中col2有null值： 如果T2中col2有null值： not in、<> all对null值敏感，即not in、、<> all后面的子查询或者…

Oracle sql 子查询

SQL Server 连接 sql server 连接方式

数据库入门～连接数据库（详细步骤+登录注册案例+简单界面）步骤一：SQL Server使用sql server身份验证登录，方便与编写…

SQL Server 连接 sql java jdbc SQL

Android 反注入调试 android sql注入

文章目录一：什么是sql注入二：SQL注入攻击的总体思路三：SQL注入攻击实例四：如何防御SQL注入1、检查变量数据类型和…

Android 反注入调试 数据库 mysql sql sql注入

pl/sql developer连接mysql sql developer 连接

plsql developer是一款集成的开发系统，它主要是针对于Oracle数据库的存储进行开发，这款软件能充分的发挥出Oracle程序优…

PLSQL 配置文件 连接数据库

hive sql左连接 sql左连接用法

1、内联接（典型的联接运算，使用像 = 或 <> 之类的比较运算符）。包括相等联接和自然联接。 内联接使用比较运算符根据…

hive sql左连接 外连接 内连接 连接查询

sql server 连接mysql sql server 连接查询

这个样例，因为在ADO.net入门已经专门学了，再次进行复习 一下。主要掌握连接字串的情况。过程就是：1、引用System.Data…

sql server 连接mysql 数据库 Data Sales

sql server DAC连接 sql server连接方式

如何使用Connection对象连接数据库？对于不同的.NET数据提供者，ADO.NET采用不同的Connection对象连接数据库。这些Co…

sql server DAC连接 数据库 Server bc

SQL server 全连接 sql server连接命令

sql server 一些较常用的操作命令

SQL server 全连接 SQL Server Windows 服务器 数据库

【SQL】:内连接，自然连接

这里主要是做一下笔记，以免自己忘记了 一.自然连接 对于自然连接而言，连接两个table之后，两个table共用的属性就会合并在…

内连接 字段 技术

sql server右连接查询sql语句 sql server 连接查询

多表连接查询：通过各个表之间共同列的关联性来查询的数据。连接查询的分类：内连接：根据表中共同的列来进行匹配。表A…

连接查询 结果集 外连接

项目根目录下运行命令yarn install来安装项目依赖 项目的根目录是什么

1. MavenProjectRoot(项目根目录)src main java ——存放项目的.java文件resources ——存放项目资源文件，如spring, hibernate…

maven xml jar

深度学习超分辨大尺寸图片如何处理 深度图超分辨率重建

文章目录0 前言1 什么是图像超分辨率重建2 应用场景3 实现方法4 SRResNet算法原理5 SRCNN设计思路6 代码实现6.1 代码结…

深度学习超分辨大尺寸图片如何处理 大数据 数据分析 python 卷积

python 不等长基因序列相似度比对 python基因差异分析

基因组结构元件的可视化有多种方式，比如IGV等基因组浏览器中以track为单位的展示形式，亦或以circos为代表的圈图形式，…

python 不等长基因序列相似度比对 python绘制基因结构图 python 数据分析 ci

汉王考勤机对接java 汉王考勤机客户端

汉王考勤程序驱动软件安装1 安装目录可以自己选择，解压软件2 进入v7_setup目录，双击Setup.exe程序，开始安装；3 安装步…

汉王考勤机对接java 服务端 客户端 管理系统

linux编译hiredis动态库 linux制作动态库

linux系统下的应用编程需要系统提供的库文件，包括静态库或动态库。不管是静态库还是动态库，都是编译好的二进制文件…

linux编译hiredis动态库 静态库 动态库 库文件