

[Pages](#) / ... / [mysql](#)

mysql Cardinality

Created by 杨超, last modified on 2018 Jun 14

关于mysql的index聚集度

表结构

```
| report | CREATE TABLE `report` (  
  `id` int(11) unsigned NOT NULL AUTO_INCREMENT,  
  `fromId` int(11) unsigned NOT NULL DEFAULT '0' COMMENT '来源id',  
  `fromType` int(11) NOT NULL DEFAULT '0' COMMENT '来源类型: 0:未知 1:spider',  
  `name` varchar(256) NOT NULL DEFAULT '' COMMENT '报告名',  
  `content` text COMMENT '文件内容',  
  `contentType` varchar(32) NOT NULL DEFAULT '' COMMENT '内容分类: bp/一级市场/二级市场/招股书',  
  `fileType` varchar(32) NOT NULL DEFAULT '' COMMENT '文件类型: pdf/docx',  
  `reportSize` int(11) NOT NULL DEFAULT '0' COMMENT '文件大小 KB',  
  `status` tinyint(3) unsigned NOT NULL DEFAULT '0' COMMENT '状态',  
  `tags` text COMMENT 'tags',  
  `downloadKeywords` text COMMENT '下载词',  
  `pages` int(11) DEFAULT '0' COMMENT 'pages',  
  `downloadTimes` int(11) NOT NULL DEFAULT '0' COMMENT '下载次数',  
  `md5Hash` char(32) NOT NULL DEFAULT '' COMMENT '文件md5hash',  
  `uploadUserId` int(11) unsigned NOT NULL DEFAULT '0' COMMENT '上传人',  
  `uploadTime` datetime NOT NULL DEFAULT '1970-01-01 00:00:00' COMMENT '上传时间',  
  `verifyUserId` int(11) unsigned NOT NULL DEFAULT '0' COMMENT '审核人',  
  `verifyUserName` varchar(256) NOT NULL DEFAULT '' COMMENT '审核人名',  
  `verifyTime` datetime NOT NULL DEFAULT '1970-01-01 00:00:00' COMMENT '审核时间',  
  `url` varchar(1024) NOT NULL DEFAULT '' COMMENT '报告的原始url',  
  `attr` text COMMENT 'attr',  
  `reportTime` datetime NOT NULL DEFAULT '1970-01-01 00:00:00' COMMENT '文件创建时间',  
  `creationTime` datetime NOT NULL DEFAULT CURRENT_TIMESTAMP,  
  `updateTime` datetime DEFAULT CURRENT_TIMESTAMP ON UPDATE CURRENT_TIMESTAMP,  
  `dt_modtime` datetime DEFAULT CURRENT_TIMESTAMP ON UPDATE CURRENT_TIMESTAMP,  
  PRIMARY KEY (`id`),  
  KEY `idx_fromId` (`fromId`),  
  KEY `sub_idx_name` (`name`(32)),  
  KEY `idx_reportTime` (`reportTime`),  
  KEY `idx_uploadTime` (`uploadTime`),  
  KEY `idx_creationTime` (`creationTime`),  
  KEY `idx_updateTime` (`updateTime`),  
  KEY `idx_dt_modtime` (`dt_modtime`),  
  KEY `idx_md5hash` (`md5Hash`)  
) ENGINE=InnoDB AUTO_INCREMENT=27058 DEFAULT CHARSET=utf8 COMMENT='行研报告' |
```

记录数

```
mysql> select count(*) from report;
+-----+
| count(*) |
+-----+
|      27056 |
+-----+
1 row in set (0.02 sec)
```

index聚集度

查看一下索引聚集度 发现id的聚集度远远大于count 429672 >> 27056

```
mysql> show index from report;
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| Table | Non_unique | Key_name | Seq_in_index | Column_name | Collation | Cardinality | Sub_part | Packed | Null | Index_type | Comment | Index_comment |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| report | 0 | PRIMARY | 1 | id | A | 429672 | NULL | NULL | | BTREE | | |
| report | 1 | idx_fromId | 1 | fromId | A | 61381 | NULL | NULL | | BTREE | | |
| report | 1 | sub_idx_name | 1 | name | A | 61381 | 32 | NULL | | BTREE | | |
| report | 1 | idx_reportTime | 1 | reportTime | A | 6 | NULL | NULL | | BTREE | | |
| report | 1 | idx_uploadTime | 1 | uploadTime | A | 3941 | NULL | NULL | | BTREE | | |
| report | 1 | idx_creationTime | 1 | creationTime | A | 702 | NULL | NULL | | BTREE | | |
| report | 1 | idx_updateTime | 1 | updateTime | A | 30690 | NULL | NULL | YES | BTREE | | |
| report | 1 | idx_dt_modtime | 1 | dt_modtime | A | 30690 | NULL | NULL | YES | BTREE | | |
| report | 1 | idx_md5hash | 1 | md5hash | A | 53709 | NULL | NULL | | BTREE | | |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
9 rows in set (0.00 sec)
```

所以问题来了，为什么会有这么大的偏差？

答案在此：<https://stackoverflow.com/questions/29441260/index-on-mysql-column-has-lower-cardinality-than-count-distinct>

<https://dev.mysql.com/doc/refman/5.5/en/show-index.html>

简单说来 cardinality是一个对index 唯一数值的个数的估算值，在mysql内部是一个静态的整形变量，并不要求精确 它的最大的意义是“告诉” mysql优化器 更大的cardinality是更应该被选择的索引

Like Be the first to like this

No labels

地址：北京市朝阳区建国路86号佳兆业广场北塔6层梦想加空间601室

以太资本由艾普拉斯投资顾问(北京)有限公司运营，提供早期互联网项目的投融资对接服务

©2014-2017 以太资本 京ICP备14028208号