

KYLIN-5393 Design & Test

Fix Design

kylin4 的fix方案是通过将单引号在后端进行转义，在下压的时候将转义后的改回来，从而使得下压和击中模型查询保持语义一致。

kylin5 在操作手册上已经明确常量中间的单引号需要通过单引号转义：

因此，我们只需要在下压的时候将两个连续的单引号 `''` 转义为一个单引号 `'` 即可，这里需要注意的是：

用户输入：三个单引号 `'aa'''bbb'` 我们会转义为 `'aa'bbb'`

问题复现

查询击中模型，结果显示正确。在查询索引的时候"等价于"。

Query Editor: `1 select * from H1.CUSTOMER where C_CITY like 'Xabb'4y00000ggzmcqk''`

Query Results:

C_CUSTKEY	C_NAME	C_ADDRESS	C_CITY	C_NATION	C_REGION	C_PHONE	C_MKTSEGMENT
120	Customer#0...	MBy6qq3O	5abb'4y000...	CHINA	ASIA	18767162214	AUTOMOBILE
118	Customer#0...	MBy6qq3O	abb'4y000...	CHINA	ASIA	18767162214	AUTOMOBILE
119	Customer#0...	MBy6qq3O	abb'4y000...	CHINA	ASIA	18767162214	AUTOMOBILE

Query Statistics: 1.01 s, 3.81 KB, 3 rows, 10 rows/page.

Model Index:

- 模型索引
- 击中模型: test1115
- 击中索引: 20000000001

Result Rows: 3

Scan Bytes: 3896

Query ID: a9edef1-7...

Query Node: snoopy-gw11.k...

Hit Cache: false

Hit Snapshot: -

查询下压结果如下，与查询击中索引结果不一致。

查询编辑器 query4 query3 query2 query1 关闭全部 已保存的查询(0)

```
1 select * from H1.CUSTOMER where C.CITY like 'abb'4y00000ggzmcagk'
```

保存 清空 Limit 0 查询 Control / Command + Enter = 查询

2 0.18 s 650 B 收起详情 <<

数据 可视化 筛选 导出 CSV

c_custkey	c_name	c_address	c_city	c_nation	c_region	c_phone	c_mktsegment
116	Customer#0...	MBY6qq3O	abb4y0000...	CHINA	ASIA	18767162214	AUTOMOBILE
117	Customer#0...	MBY6qq3O	abb4y0000...	CHINA	ASIA	18767162214	AUTOMOBILE

共 2 条 10条/页 < 1 > 前往 1 页

查询诊断

下压

查询时间

0.18s

未构建模型索引。构建后可加速查询。前往构建

数据源

查询对象

HIVE

结果行数 2

扫描字节 650

查询 ID 9562015a-...

查询节点 snoopy-gw11k...

击中缓存 false

击中快照

Test

准备数据

```
1 CREATE TABLE test_symbol(  
2   id bigint,  
3   trans_id bigint,  
4   order_id bigint,  
5   cal_dt date,  
6   cal_dt_name string,  
7   cal_dtime string,  
8   cal_ts timestamp)  
9 ROW FORMAT DELIMITED  
10  FIELDS TERMINATED BY ','  
11 STORED AS INPUTFORMAT  
12  'org.apache.hadoop.mapred.TextInputFormat'  
13 OUTPUTFORMAT  
14  'org.apache.hadoop.hive ql.io.HiveIgnoreKeyTextOutputFormat'  
15 TBLPROPERTIES (  
16  'COLUMN_STATS_ACCURATE'='false',  
17  'last_modified_by'='root',  
18  'last_modified_time'='1647222536',  
19  'numFiles'='1',  
20  'numRows'='-1',  
21  'rawDataSize'='-1',  
22  'totalSize'='1093073',  
23  'transient_lastDdlTime'='1647222536');  
24  
25 insert into test_symbol(id,trans_id,order_id,cal_dt,cal_dt_name,cal_dtime,cal_
```

```

26 (1,1,101,'2022-05-01','20220'101','2022-05-01 00:01:00','2022-09-20 10:13:01')
27 ,(2,1,102,'2022-05-02','20220'202','2022-05-01 00:02:00','2022-09-20 10:13:02')
28 ,(3,1,103,'2022-05-03','20220303','2022-05-01 00:03:00','2022-09-20 10:13:03')
29 ,(4,4,104,'2022-05-04','20220404','2022-05-01 00:04:00','2022-09-20 10:13:04')
30 ,(5,4,105,'2022-05-05','20220505','2022-05-01 00:05:00','2022-09-20 10:13:05')
31 ,(6,6,106,'2022-05-06','20220606','2022-05-01 00:06:00','2022-09-20 10:13:06')
32 ,(7,7,107,'2022-05-07','20220707','2022-05-01 00:07:00','2022-09-20 10:13:07')
33 ,(8,8,108,'2022-05-08','20220808','2022-05-01 00:08:00','2022-09-20 10:13:08');

```

测试sql:

```
1 select CAL_DT_NAME from TEST_SYMBOL where CAL_DT_NAME like '%''%'
```

预期：查询带两个单引号的字符串的sql，命中索引和下压的时候结果保持一致：

```

OK
Time taken: 21.766 seconds
hive>
>
>
> select * from test_symbol;
OK
1      1      101      2022-05-01      20220'101      2022-05-01 00:01:00      2022-09-20 10:13:01
2      1      102      2022-05-02      20220'202      2022-05-01 00:02:00      2022-09-20 10:13:02
3      1      103      2022-05-03      20220303      2022-05-01 00:03:00      2022-09-20 10:13:03
4      4      104      2022-05-04      20220404      2022-05-01 00:04:00      2022-09-20 10:13:04
5      4      105      2022-05-05      20220505      2022-05-01 00:05:00      2022-09-20 10:13:05
6      6      106      2022-05-06      20220606      2022-05-01 00:06:00      2022-09-20 10:13:06
7      7      107      2022-05-07      20220707      2022-05-01 00:07:00      2022-09-20 10:13:07
8      8      108      2022-05-08      20220808      2022-05-01 00:08:00      2022-09-20 10:13:08
Time taken: 0.173 seconds, Fetched: 8 row(s)
hive>

```

查询编辑器query1 ×关闭全部已保存的查询(0)

1 select CAL_DT_NAME from TEST_SYMBOL where CAL_DT_NAME like 'X'%'

保存 清空Limit 500 查询

Control / Command + Enter = 查询

查询结果

查询 ID: cebd5773-2552-e955-5a81-cd4a8c828b05

查询对象: test_model_01

索引 ID: 20000000001

查询节点: sandbox.hortonworks.com:17072

查询耗时: 4.01s

查询扫描记录数: 8

查询结果记录数: 2

数据 可视化

Q 筛选

导出 CSV

CAL_DT_NAME

20220'101

20220'202

共 2 条10条/页 < 1 > 前往 1 页

查询编辑器query1 ×关闭全部已保存的查询(0)

1 select CAL_DT_NAME from TEST_SYMBOL where CAL_DT_NAME like 'X'%'

保存 清空Limit 500 查询

Control / Command + Enter = 查询

查询结果

查询 ID: 77ad7cd9-7b92-8d46-2ef0-0db#3cb3aa2

查询对象: HIVE

查询节点: 192.168.101.9:7070

查询耗时: 81.05s

数据 可视化

Q 筛选

导出 CSV

CAL_DT_NAME

20220'101

20220'202

共 2 条10条/页 < 1 > 前往 1 页