



—— 数据库年终盘点大会-上海站 ——

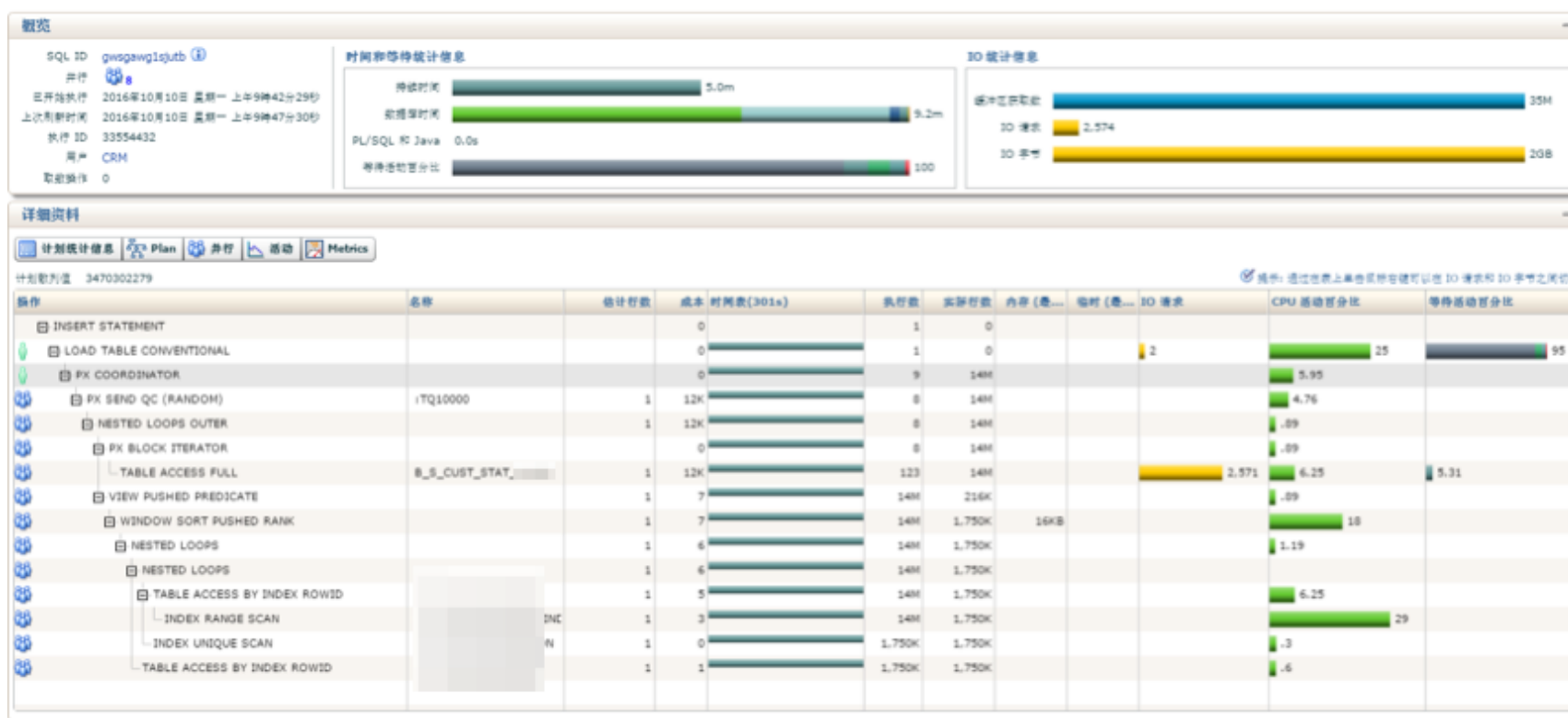
Oracle 动态采样

蒋健

Oracle 动态采样经典场景

- SQL使用业务场景临时表
- SQL组合过滤条件相对复杂
- 复杂聚合

临时表和动态采样



表B_S_CUST_STAT的过滤条件为Data_date = to_date(20161008,'yyyymmdd'), 估算值同样为一行, 实际值为一千四百万行, 导致后续连接方式为 nested loop outer, 被驱动视图表访问了一千四百万次

临时表和动态采样

B_S_CUST_STAT_

Object Type:

Table

Owner:

CRM

Object Name:

B S CUST STAT

[查看数据库对象](#)

(大小写敏感)

表

列

卷三

約束

分区

子分区

OWNER	TABLE_NAME	NUM_ROWS	BLOCKS	AVG_ROW_LEN	GLOBAL_STATS	SAMPLE_SIZE	LAST_ANALYZED
CRM	B_S_CUST_STAT_I	13799310	316380	162	YES	1379931	2016-10-10 08

B_S_CUST_STAT_

Object Type:

Table

Owner

CRM

Object Name:

B S CUST STAT

[查看数据库对象](#)

(大小写敏感)

表

列

索引

约束

分区

子分区

[illegible]

临时表和动态采样

```
create or replace function raw_to_date(i_raw raw)
return date
as
m_n date;
begin
dbms_stats.convert_raw_value(i_raw,m_n);
return m_n;
end;
/

select raw_to_date('78740A07010101') stats_value from dual;

STATS_VALUE
-----
2016-10-07 00:00:00
```

Data_date 列上只有一个唯一值，为10月7号。对表 P_CUST_STAT的使用 Data_date = to_date(20161008,'yyyymmdd') 的查询的估算值都为1。

临时表和动态采样

删除表B_S_CUST_STAT和P_CUST_STAT的统计信息并且进行锁定，保证后续对临时表的查询会使用动态采样，得到准确的估算值。

```
Exec dbms_stats.delete_table_stats('CRM','P_CUST_STAT');  
Exec dbms_stats.lock_table_stats('CRM','P_CUST_STAT');
```

复杂查询与动态采样

对于有复杂的过滤条件的sql, 为了在执行计划中得到正确的cardinality, 统计信息未必有帮助, 包括extended statistics. 比如in和like的组合条件, 或者where条件中使用了自定义的函数. 这时候dynamic sampling可能是唯一的选择. 下面是一个例子, 采用level为6的采样之后, cardinality更为接近真实的数据.

复杂查询与动态采样

```
create table t1 as
with v1 as (
select /*+ materialize */
rownum id from dual connect by level <= 1000
)
select
    rownum                id,
    rpad(rownum, 10, '0')  v1,
    trunc((rownum - 1)/100) n1,
    case
        when mod(rownum,100000) = 7          then 'ERR'
        when rownum <= 9990000               then 'COM'
        when mod(rownum,10) =0                then 'NEW'
        when mod(rownum,10) between 1 and 5   then 'PRP'
        when mod(rownum,10) between 6 and 8   then 'FKC'
        when mod(rownum,10) = 9               then 'LDD'
    end status,
    rpad(rownum, 100)      padding
from v1, v1
where rownum <= 1e6;

begin
    dbms_stats.gather_table_stats(user,'t1');
end;
/
```


复杂查询与动态采样

```
1  select
2      count(*)
3  from
4      t1
5  where
6      status in ('COM', 'ERR')
7* and v1 like '10%'
SQL> /
```

```
      COUNT(*)
-----
      11113
```

Execution Plan

Plan hash value: 4096694858

Id	Operation	Name	Rows	Bytes	Cost (%CPU)	Time
0	SELECT STATEMENT		1	15	4983 (1)	00:00:01
1	SORT AGGREGATE		1	15		
* 2	TABLE ACCESS STORAGE FULL	t1	395	5925	4983 (1)	00:00:01

Predicate Information (identified by operation id):

```
2 - storage("V1" LIKE '10%' AND ("STATUS"='COM' OR "STATUS"='ERR'))
    filter("V1" LIKE '10%' AND ("STATUS"='COM' OR "STATUS"='ERR'))
```

复杂查询与动态采样

```
SQL> select /*+ OPT_PARAM('OPTIMIZER_DYNAMIC_SAMPLING', 6) */
2      count(*)
3  from
4      t1
5  where
6      status in ('COM', 'ERR')
7  and v1 like '10%';
```

COUNT(*)

11113

Execution Plan

Plan hash value: 4096694858

Id	Operation	Name	Rows	Bytes	Cost (%CPU)	Time
0	SELECT STATEMENT		1	15	4983 (1)	00:00:01
1	SORT AGGREGATE		1	15		
* 2	TABLE ACCESS STORAGE FULL	t1	16595	243K	4983 (1)	00:00:01

Predicate Information (identified by operation id):

2 - storage("V1" LIKE '10%' AND ("STATUS"='COM' OR "STATUS"='ERR'))
filter("V1" LIKE '10%' AND ("STATUS"='COM' OR "STATUS"='ERR'))

Note

- dynamic statistics used: dynamic sampling (level=6)

使用动态采样，级别为6，估算值为16595行，实际为11113行，差距不到2倍，估算值的质量大幅提升

复杂聚合

select /*+ OPT_PARAM('OPTIMIZER_DYNAMIC_SAMPLING',6) */ mod(t1.v1,13) from t1 where
t1.v1 like '10%' group by mod(t1.v1,13);

```
SQL> select /*+ OPT_PARAM('OPTIMIZER_DYNAMIC_SAMPLING',6) */ mod(t1.v1,13) from t1 where
```

```
13 rows selected.
```

```
Execution Plan
```

```
-----  
Plan hash value: 136660032
```

Id	Operation	Name	Rows	Bytes	Cost (%CPU)	Time		

0	SELECT STATEMENT		10002	107K	4936 (1)	00:00:01		
1	HASH GROUP BY		10002	107K	4936 (1)	00:00:01		
* 2	TABLE ACCESS FULL	T1	11122	119K	4935 (1)	00:00:01		

```
Predicate Information (identified by operation id):  
-----
```

12c group by与动态采样

```
SQL> select /*+ OPT_PARAM('OPTIMIZER_DYNAMIC_SAMPLING',11) */ mod(t1.v1,13) from t1 where
```

13 rows selected.

Execution Plan

Plan hash value: 136660032

Id	Operation	Name	Rows	Bytes	Cost (%CPU)	Time
0	SELECT STATEMENT		12	285	4940 (1)	00:00:01
1	HASH GROUP BY		12	285	4940 (1)	00:00:01
* 2	TABLE ACCESS FULL	T1	19244	413K	4938 (1)	00:00:01

Predicate Information (identified by operation id):

2 - filter("T1"."V1" LIKE '10%')

Note

- dynamic statistics used: dynamic sampling (level=AUTO)

The logo for DBAplus, featuring the letters 'DBA' in red, blue, and orange respectively, followed by 'plus' in green. A thin white horizontal line is positioned below the letters.

DBAplus

www.dbaplus.cn

THANK YOU!