

COOL data overhead

For a folder with 150 days of data (SV MC example) here are the statistics :

COLUMN_NAME	AVG_COL_LEN
CHANNEL_ID	4
IOV_SINCE	8
IOV_UNTIL	8
LASTMOD_DATE	34
NEW_HEAD_ID	2
OBJECT_ID	6
ORIGINAL_ID	2
S	101
SYS_INSTIME	34
USER_TAG_ID	2

S is the payload.

```
SQL> exec DBMS_STATS.gather_table_stats('ST_CLIENT', 'PDBST005_F0001_IOVS', NULL, 100);
```

```
SQL> select avg_row_len from user_tables where table_name = 'PDBST005_F0001_IOVS';
```

```
AVG_ROW_LEN
-----
          198
```

Actually, using the user_extents table we find another result :

```
SQL> select sum(bytes) from user_extents where SEGMENT_NAME like 'PDBST005_F0001_IOVS';
```

```
SUM(BYTES)
-----
2080374784
```

```
SQL> select num_rows from user_tables where table_name = 'PDBST005_F0001_IOVS';
```

```
NUM_ROWS
-----
      8640200
```

And $2080374784/8640200 = 240\text{Bytes/row}$

For the indexes :

```
SQL> select sum(bytes) from user_extents where SEGMENT_NAME like 'PDBST005_F0001_IOVS%INDX' OR S
```

```
SUM(BYTES)
-----
      947978240
```

$947978240/8640200 = 109\text{Bytes/row}$

So in total 349Bytes/row meaning a factor of 3.5 for a payload of 100Bytes.

In the current COOL schema, the string (VARCHAR2(255)) used to represent a date takes 33 Bytes.

```
SQL> select SYS_INSTIME, length(SYS_INSTIME) from PDBST005_F0001_IOVS where rownum < 2;
```

```
SYS_INSTIME                                LENGTH(SYS_INSTIME)
```

CoolDataOverhead < PSSGroup < TWiki

2007-06-09_05:12:26.880713000 GMT 33

SQL> desc PDBST005_F0001_IOVS

Name	Null?	Type
OBJECT_ID	NOT NULL	NUMBER(10)
CHANNEL_ID		NUMBER(10)
IOV_SINCE		NUMBER(20)
IOV_UNTIL		NUMBER(20)
USER_TAG_ID		NUMBER(10)
SYS_INSTIME		VARCHAR2(255)
LASTMOD_DATE		VARCHAR2(255)
ORIGINAL_ID		NUMBER(10)
NEW_HEAD_ID		NUMBER(10)
S		VARCHAR2(4000)

From the Oracle documentation : Oracle Database stores dates in its own internal format. Date data is stored in fixed-length fields of seven bytes each, corresponding to century, year, month, day, hour, minute, and second.

So we would save 52 Bytes per row by switching to Oracle date.

After moving one of the tables to an "enhanced" version with :

```
create table PDBST005_F0001_IOVS_switch as
(
select  OBJECT_ID,
        CHANNEL_ID,
        IOV_SINCE,
        IOV_UNTIL,
        USER_TAG_ID,
        TO_DATE(SUBSTR(SYS_INSTIME,0,19), 'YYYY-MM-DD_HH24:MI:SS') "SYS_INSTIME",
        TO_DATE(SUBSTR(LASTMOD_DATE,0,19), 'YYYY-MM-DD_HH24:MI:SS') "LASTMOD_DATE",
        ORIGINAL_ID,
        NEW_HEAD_ID,
        S
from    PDBST005_F0001_IOVS
);
```

And the result in term of space is the following :

```
SQL> select sum(bytes)/power(2,30)||' GB' "Size" from user_extents where segment_name = 'PDBST005
```

Size

1.9375 GB

```
SQL> select sum(bytes)/power(2,30)||' GB' "Size" from user_extents where segment_name = 'PDBST005
```

Size

1.4375 GB

It should not affects indexes because none of them is based on the modified columns.

Here there was a 100 Bytes payload and we saved around 18% of space (taking into account the indexes).

Actually, the current precision is a fraction of a second so we should use the **TIMESTAMP** datatype which takes from 7 to 11 Bytes depending on the precision.

CoolDataOverhead < PSSGroup < TWiki

(see : http://download-uk.oracle.com/docs/cd/B19306_01/server.102/b14200/sql_elements001.htm#i54330)

With the full precision (11 Bytes) we still would save 44 Bytes per IOV.

This topic: PSSGroup > CoolDataOverhead

Topic revision: r5 - 2007-07-25 - RomainBasset



Copyright &© 2008-2021 by the contributing authors. All material on this collaboration platform is the property of the contributing authors.
or Ideas, requests, problems regarding TWiki? use Discourse or Send feedback