

DM

Data Life Cycle Panel COMPASS

Distributed Database Operations Workshop

November 11th, 2008

Dawid Wójcik, CERN / IT-DM



- COMPASS
 - fixed-target experiment at CERN studying the structure of the nucleon and spectroscopy.
 - 500 TB of data (during 2002 and 2003 runs), estimated of 300 TB data/year.
 - At the beginning these data together with the reconstructed events information were put in CASTOR and metadata in a database infrastructure based on Objectivity/DB.
 - Starting from 2003 Oracle has been adopted as the database technology for storing experiment metadata (currently over 7.2 TB of data in Oracle DB).

- Event metadata organized by weeks into separate users with different tablespaces
- Each user have the same tables' structure
- IOTs (Index Organized Tables) used to store data and retrieve them very efficiently
- One of the biggest schema (event metadata of week 39/2008) has ~220GB of data (over 4 billion rows in one table!)

The screenshot displays the Oracle SQL Developer interface. The left pane shows a tree view of the database schema, including users (COMPASS_08W26 to COMPASS_08W39), tables (EVENT_HEADERS, FILE_MAPS, LOCKTAB, RUNS, SPCEV_SEGMENTS, WRONG_EVENT_HEADERS), views, indexes, packages, and procedures. The right pane shows the structure of the selected table, COMPASS_08W39.EVENT_HEADERS, with columns and their properties.

Column Name	Data Type	Nullable	Data Default	COLUMN ID	Primary Key	COMMENTS
RAWEV_FILE_ID	NUMBER(8,0)	No	(null)	10	1 (null)	1 (null)
EVENT_NUMBER	NUMBER(10,0)	No	(null)	2	2 (null)	2 (null)
RUN_NUMBER	NUMBER(8,0)	No	(null)	1	(null) (null)	(null) (null)
BURST	NUMBER(9,0)	No	(null)	3	(null) (null)	(null) (null)
EVENT_IN_BURST	NUMBER(8,0)	No	(null)	4	(null) (null)	(null) (null)
TRIGGER_MASK	NUMBER(12,0)	No	(null)	5	(null) (null)	(null) (null)
TIME_SEC	NUMBER(12,0)	No	(null)	6	(null) (null)	(null) (null)
TIME_USEC	NUMBER(12,0)	No	(null)	7	(null) (null)	(null) (null)
ERROR_CODE	NUMBER(12,0)	No	(null)	8	(null) (null)	(null) (null)
EVENT_SIZE	NUMBER(8,0)	No	(null)	9	(null) (null)	(null) (null)
RAWEV_FILE_OFFSET	NUMBER(12,0)	No	(null)	11	(null) (null)	(null) (null)

Oracle SQL Developer : TABLE COMPASS_08W39.EVENT_HEADERS@compr
All Rows Fetched: 11 | compr | COMPASS_08W39 | EVENT_HEADERS Editing

- Reconstruction metadata organized into separate users with different tablespaces
- Each user have the same tables' structure
- IOTs (Index Organized Tables) used to store data and retrieve them very efficiently
- One of the biggest schema has ~139GB of data (over 3 billion rows in one table!)

The screenshot displays the Oracle SQL Developer interface. On the left, a tree view shows the database structure for user COMPDST_03P1J, including tables like COMPDST_02P2H, COMPDST_03P1A through COMPDST_03P1J, and a tablespace named DST. The main window shows a schema diagram with the following tables and their attributes:

- RUN**: # run number, o time, o status, o logbook
- RAW FILE**: # file ID, u file name
- DST FILE**: # file ID, u file name, * DST version, * DST type, o value1 descr, o value2 descr, o value3 descr
- EVENT HDR**: # event number, * event size, * event filepos, * burst number, * event in burst, * trigger mask, * time, * error code
- DST HEADER**: # event number, * DST size, * DST filepos, * trigger mask, o value1, o value2, o value3

Relationships are shown with lines connecting the tables. A table view for 'EVENT_FLAG' is also visible, showing columns and their properties:

COLUMN ID	Primary Key	COMMENTS
1	1 (null)	
2	2 (null)	
3	(null) (null)	
4	(null) (null)	
5	(null) (null)	
6	(null) (null)	
7	(null) (null)	
8	(null) (null)	
9	(null) (null)	

The bottom status bar indicates 'All Rows Fetched: 9' and the current session is 'compr | COMPDST_03P1J | DST Editing'.

- Advantages
 - Data fully separated – self contained set of data (different schemas/tablespaces)
 - Manageability – data can be easily exported/imported/dropped (schema export/datapump)
- Disadvantages
 - Complex cross-schema dependencies may be needed to be maintained (not an issue for COMPASS)
 - Password and privileges management issues (multiple schemas for single application) – higher management overhead
 - Common reader/writer accounts and access management needed

DM

```
// for all nodes,  
for(tp = m; tp < m + n; tp++)  
if(tp->second->busyTPools.p  
// reap child pr  
pid_t pid;  
while ((pid = w  
if(!beGraceful)  
// on a SIGINT  
return; //  
}  
// now loop wait  
while(busyTPool  
sleep(1); // v  
for(unsigned i  
if(busyTPools  
// it's idle no  
busyTPools.  
else
```

Q & A

