CERN IT Department

Data Life Cycle Panel
COMPASS

Distributed Database Operations Workshop

November 11th, 2008

Dawid Wójcik, CERN / IT-DM

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

COMPASS experiment

CERN IT Department

- 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).

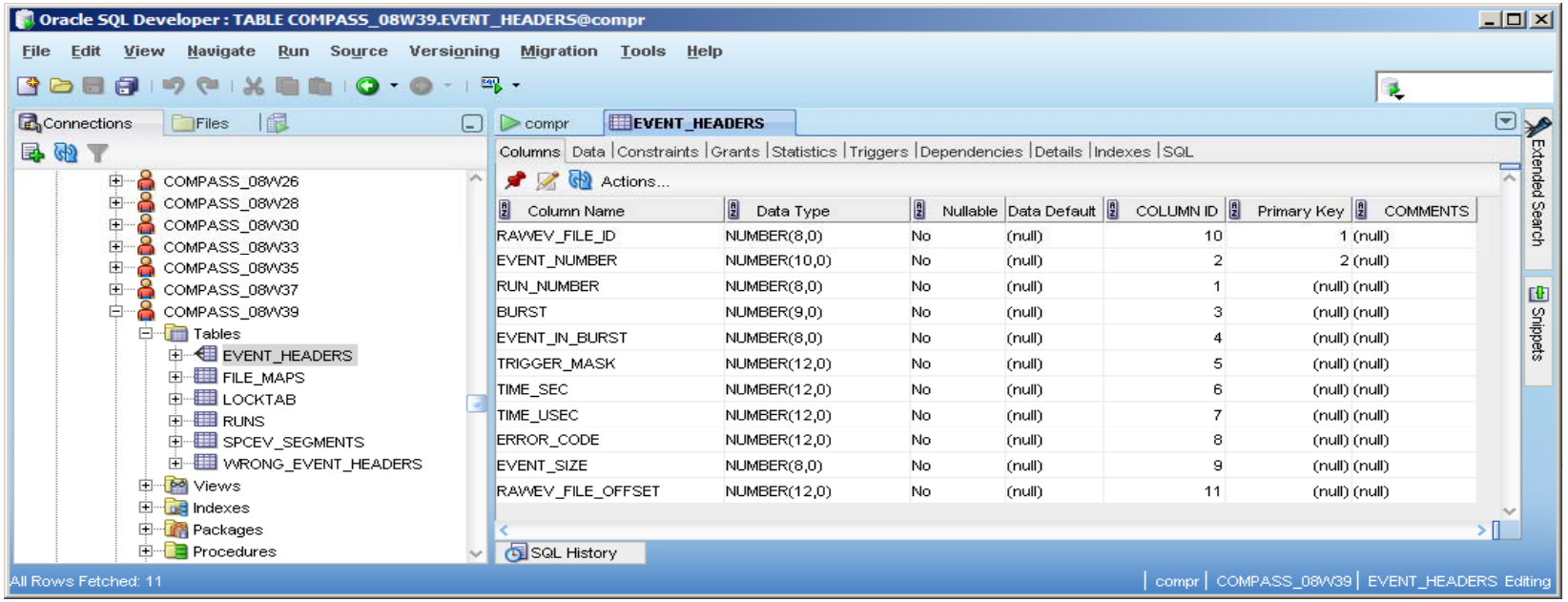
CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

CERN

DM COMPASS event metadata

CERN IT Department

- 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!)

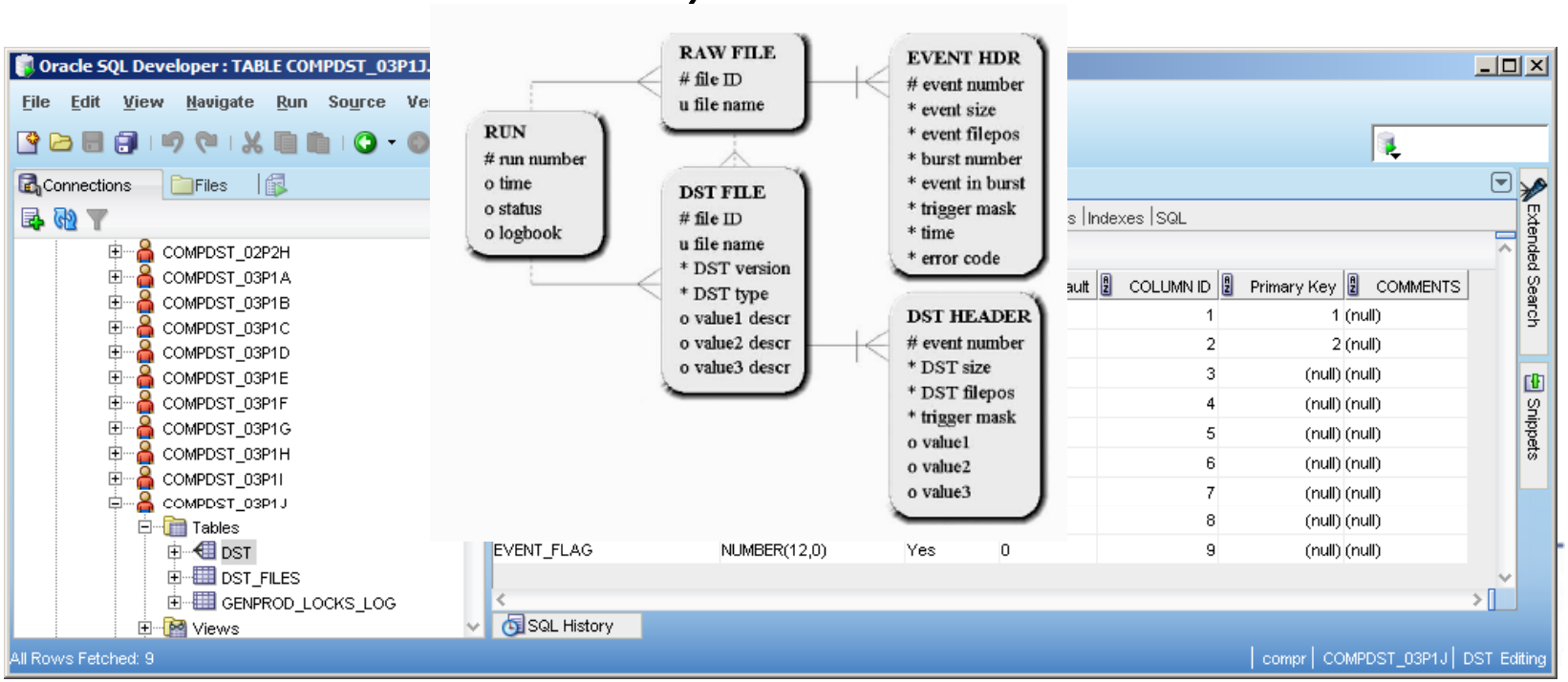


CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

DM COMPASS reconstruction metadata

CERN IT Department

- 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!)



CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

DM COMPASS – Data Life Cycle

CERN IT Department

- 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

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

CERN

Q & A

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it