Oracle Advanced Compression Tests

Svetozar Kapusta
15th of October 2009
What is CERN?

- CERN is the world’s largest particle physics laboratory located in Geneva, Switzerland
- CERN hosts the Large Hadron Collider (LHC) which is the biggest man-made accelerator

- LHC will start its operation in November 2009 and will form, together with its experiments, the biggest sub-nuclear microscope in the world.

CERN is:
- ≈2500 staff scientists (physicists, engineers, etc.)
- ≈6500 visiting scientists (half of the world's particle physicists)
- Coming from ≈500 universities or institutes representing ≈80 nationalities.

Courtesy of M. Girone
LHC: a Very Large Scientific Instrument

- LHC: 27 km long
- 100m underground

ALICE
ATLAS
CMS
+TOTEM

LHCb

Mont Blanc, 4810 m

Downtown Geneva

Courtesy of M. Girone
Based on Advanced Technology

27 km of superconducting magnets cooled in superfluid helium at 1.9 K

Courtesy of M. Girone
Experiments are ready for collisions

Courtesy of M. Girone
~ 300,000 MB/s
from all sub-detectors

~ 300MB/s
Raw Data

Trigger and data acquisition

Event filter computer farm

Courtesy of M. Girone
Data Acquisition, First pass processing

LHCb ~ 50 MB/sec
ATLAS ~ 320 MB/sec
ALICE ~ 100 MB/sec
CMS ~ 220 MB/sec
1.25 GB/sec (ions)

Courtesy of M. Girone
CERN Openlab

- Collaboration between CERN and industrial Openlab partners: HP, Intel, Oracle and Siemens
- Framework for evaluating and integrating cutting-edge IT technologies
- CERN acquires early access to technology
- CERN offers expertise and a demanding computing environment to push new technologies to their limits
- CERN provides a neutral ground for carrying out advanced R&D
- Excellent collaboration with Oracle
Relational databases play a key role in the experiments’ production dataflow chains

Listed among the critical services for the LHC experiments

Bulk of physics data stored in files, a fraction of it in databases

Most applications are OLTP

Some data warehouse applications are also emerging
Data Growth

- Expected data growth is roughly \( \approx 20-30 \ TB \) per year per experiment
- Experiments need to have all data available at any time
  - During the experiments lifetimes (10-15 years)
  - Few extra years, as the data analysis will continue
- We have to provide an efficient way of storing and accessing the few Peta bytes of mostly read-only data
- Answer to our challenge is the compression available in 11G2 and Exadata2
Advanced Compression Tests

- Exadata2 located in Reading, UK
  - Half rack with 7 storage cells each of 12 disks each
  - Accessed remotely from Geneva for 2 weeks

- Data used
  - The largest and representative production and test tables
  - Exported compressed using Datapump
  - Imported into Exadata2 using Datapump

- Applications
  - PVSS (slow control system used by the experiments)
  - GRID monitoring application
  - GRID Test data
  - File transfer applications (PANDA)
  - Logging application for ATLAS

- First results the same day
Compression factors for various compression types of various physics applications

PVSS columns: 6 number, 4 TS(9), 5 varchar2, 3 binary_double
LCG GRID Monitoring columns: 5 number
LCG TESTDATA columns: 6 number(38), 1 varchar2, 1 CLOB
ATLAS PANDA FILESTABLE columns: 3 number, 12 varchar2, 2 date, 2 char
ATLAS LOG MESSAGES columns: 5 number, 7 varchar2, 1 TS
Table creation times for various compression types of various physics applications. Normalized to no compression.

PVSS columns: 6 number, 4 TS(9), 5 varchar2, 3 binary_double
LCG GRID monitoring columns: 5 number
LCG TESTDATA columns: 6 number(38), 1 varchar2, 1 CLOB
ATLAS PANDA FILESTABLE columns: 3 number, 12 varchar2, 2 date, 2 char
ATLAS LOG MESSAGES columns: 5 number, 7 varchar2, 1 TS
Full table scans performance for various compression types of various physics applications. Normalized to no compression.

- PVSS columns: 6 number, 4 TS(9), 5 varchar2, 3 binary_double
- LCG GRID monitoring columns: 5 number
- LCG TESTDATA columns: 6 number(38), 1 varchar2, 1 CLOB
- ATLAS PANDA FILESTABLE columns: 3 number, 12 varchar2, 2 date, 2 char
- ATLAS LOG MESSAGES columns: 5 number, 7 varchar2, 1 TS
Full table scans performance for various compression types of various physics applications. Normalized to no compression. Exadata offloading set to false.

- PVSS columns: 6 number, 4 TS(9), 5 varchar2, 3 binary_double
- LCG GRID monitoring columns: 5 number
- LCG TESTDATA columns: 6 number(38), 1 varchar2, 1 CLOB
- ATLAS PANDA FILESTABLE columns: 3 number, 12 varchar2, 2 date, 2 char
- ATLAS LOG MESSAGES columns: 5 number, 7 varchar2, 1 TS
Exadata2 offloading

Full table scans performance for various compression types of ATLAS logging application with and without Exadata offloading

Please note the logarithmic scale
Export Datapump Compression

- Compression factor for PVSS data
  - Export Datapump ≈ 9X
  - *tar bzip2 utility*
    - ≈ 11X on non compressed exported PVSS data
    - ≈ 1.2X on the compressed exported PVSS data

- Compression factor for LCG application
  - Export Datapump ≈ 13X
  - *tar bzip2 utility*
    - ≈ 9X on non compressed exported LCG data
    - ≈ 1.2X on the compressed exported LCG data
Conclusions

- Tested basic, OLTP and hybrid columnar compression and Datapump compression
- The results for data from physics applications are rather impressing (2-6X OLTP, 10-70X EHCC archive high)
- EHCC can achieve up to \( \approx 3X \) better compression than tar bzip2 compression of the same data exported uncompressed
- Oracle Compression offers a win-win solution, especially for OLTP
  - Shrinks used storage volume
  - Improves performance
Thank you for your attention