LCG Gridview / LCG SAM use cases

Miguel Anjo

8th July 2008

Database Developers' Workshop

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

Scenario

CERN IT Department

- Summarization and Monitoring of Grid Sites and Services
- Large schemas (>400GB each)
- 30GB new data/month
- Big raw data tables (>300M rows in Gridview case)

- Need to partition tables
  - Used RANGE partition on a timestamp column

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

CERN

LCG Grid View example

CERN IT Department

- RAW results table
  - Monthly partitions
  - Drop old partitions after three months
- Summarized table
  - Big composite (multi-column) indexes to help performance of aggregations
  - Drawback: space used by indexes >= space used by table
- Big merge queries
  - Not easy to tune
  - Working to transform in PL/SQL bulk operations

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

LCG SAM example

CERN IT Department

- Raw data table
  - includes CLOB column with details
  - CLOB can't be easily deleted
  - Interest on keep >1 year raw data (except CLOB)
  - But no access to data >3 months
- Current solution
  - Partitioned raw data table
  - Created monthly partitioned "History" table
    - No indexes, no CLOB column
  - Move data >3 months to "History" table
  - Gain of 80% space per month

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

Lessons learned

CERN IT Department

- Necessary to think on the growth
- Partition is not (yet) self managed
  - Necessary to manually drop/create partitions
  - User responsibility to manage partitions
- Operations with partitions are bit tricky
  - But difficult to test with few data
- Still not good solution for keeping old, not accessed, data
  - User would like a DVD with it ☺
  - History table, no indexes, read only tablespace

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

CERN