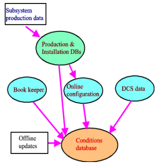
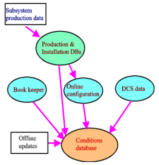
Database TEG workshop, Nov 2011



ATLAS Oracle database applications and plans for use of the Oracle 11g enhancements

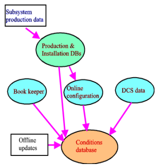
Gancho Dimitrov



Outline

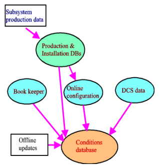
- Some facts about the ATLAS databases at CERN
- Plan for upgrade to 11g version plus hardware migration
- The current operational challenges
- How to address the challenges?
- Key messages

The ATLAS databases at CERN



Database roles, current Oracle versions, plans for upgrade to Oracle 11g

Database/ fact	ATONR	ATLR	ADCR	ATLARC	INTR	INT8R
Main database role	Online data taking	Post data taking analysis	Grid jobs and file management	Event metadata (TAGS)	SW and replica tests	SW tests
Current Oracle ver.	10.2.0.5	10.2.0.5	10.2.0.5	10.2.0.5	10.2.0.5	11.2.0.3
Planned for upgrade to 11g ver.	Jan. 2012	Jan 2012	Jan. 2012	End of November 2011		

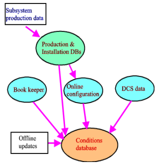


The ATLAS databases at CERN (facts)

The snapshot represents the current state with an **impressive number of database schemas.** In the common case each schema is related to a dedicated client application except the COOL, PVSS and PVSSCONF schemas as then each account maps to a sub-detector.

Database / Metric	ATONR	ATLR	ADCR	ATLARC
# DB schemes	69	151	12	51 (incl 23 arch)
# Tables - Non partitioned - Partitioned	 37045 384	 51019 665	 364 21	 9033 6886
Volume - Table segments - Index segments - LOB segments	 2,14 TB 2,15 TB 0,3 TB	 4.6 TB 6,2 TB 0,4 TB	 5,4 TB 2,1 TB (incl. IOTs) 0,3 TB	 4,2 TB 10,7 TB -
Average daily segments growth (year 2011)	8.5 GB/day	14.2 GB/day	9 GB/day	24.2 GB/day

The current operational challenges



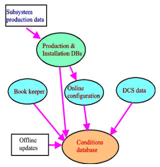
- **1) Scalability**

  => Are we able to scale on the DB side with the increased user workload?

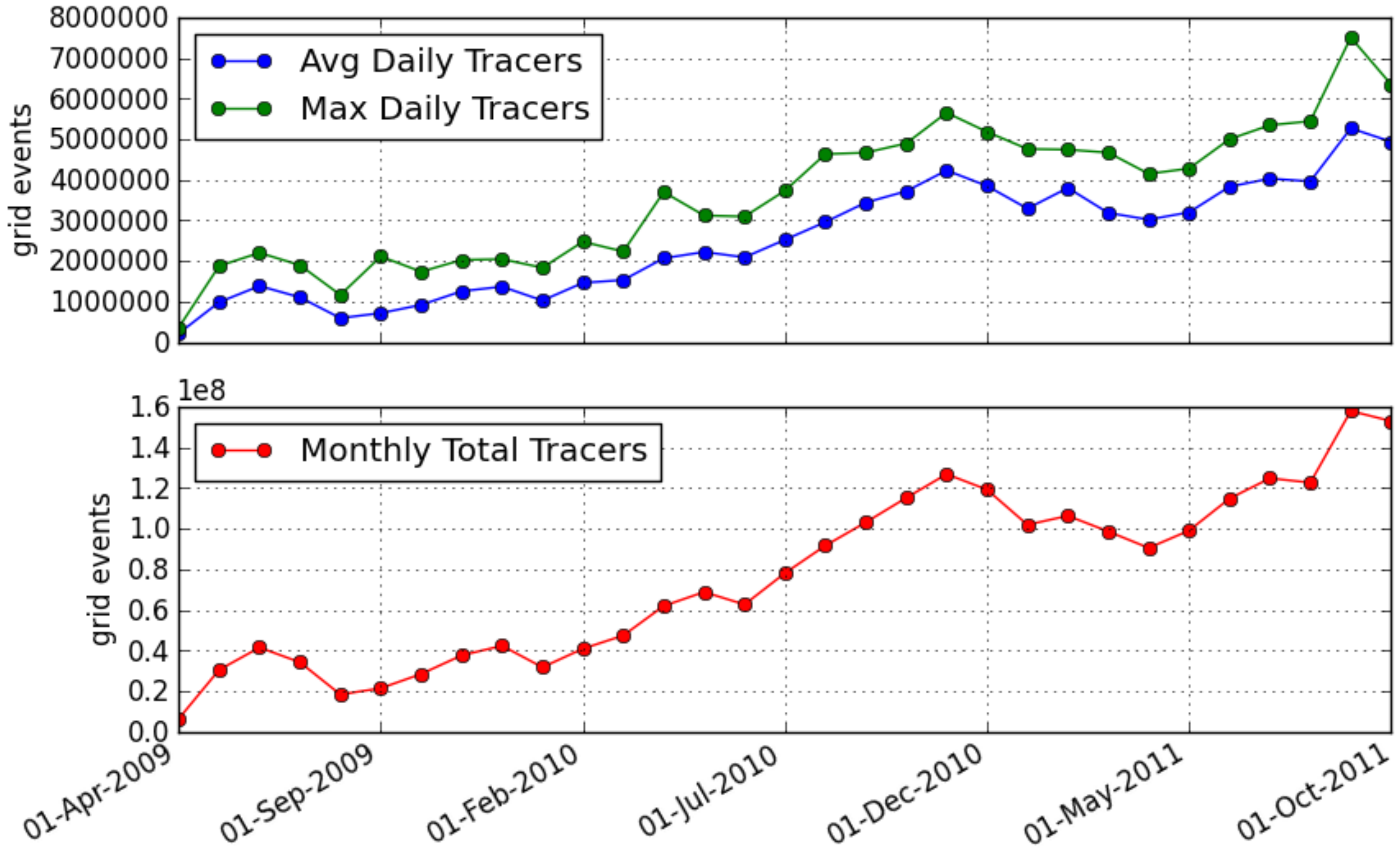
  e.g. the activity on the grid and available resources is increasing with the time (the plot on the next slide) which puts more and more load on the ATLAS DDM (Data Management and Distribution) system and the Oracle database backend.

  We, the DBAs and developers, do a lot of SQL and PLSQL tuning to make the system more efficient, but in the last weeks we were operating close to the ADCR database HW limits!

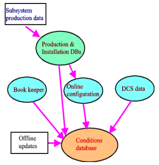
  Can the activity on the grid be controlled/contrained?



The current operational challenges (2)



The current operational challenges (3)



- **2) User’s workload**

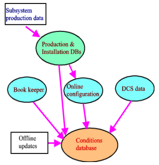
  => Important question is what fraction of the users load is legitimate?

  Examples: obsolete crons that pull data, execution of queries with high rate and probably for the same bind variable values.

- **3) The hardware**

  => The HW of all ATLAS databases is with the same characteristics because of cost, maintenance and operational reasons, but the challenge is that they have to serve different workloads!

The current operational challenges (cont.)



- **4) Database workload type**

  => the ATONR is strictly of OLTP workload type (for transaction-oriented applications) with a sliding window of an year for PVSS data imposed from the DBAs as agreed (equivalent to about 3 TB)

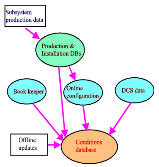
  => ATLARC is of type data warehouse.

  => the ATLR and ADCR hosts many versatile applications and that is why is a kind of a mixture between OLTP and data warehouse. For that reason these databases are the most difficult to deal with.

- **5) Users' needs**

  => Users want consistent data in the database, but also want free style searches resolved within few seconds for applications that are now hosted on the ATLR and ADCR databases.

  Current issue: Users' wildcard searches are translated from Oracle in full index or full table scans. For the moment it is not feasible to keep the relevant data segments always in the server's cache (to avoid the disk reads) because of the lack of enougn memory.



To address the challenges ahead ...

- **1) Stronger hardware (scale vertically or horizontally?)**

  => As of 2012 stronger hardware (faster CPUs and 3 times more memory) will replace the current one which will give us some room, but is not a full remedy. Scaling horizontally is not appropriate for applications with high transaction rate

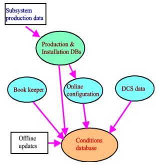
  (e.g. the ATLAS DDM system (Data Management and Distribution) transaction rate is in the range 800-1200 Hz. The value of that metric for PanDA (Production ANd Distributed Analysis) is 150-200)

- **2) Oracle 11g new features and enchancements**

  (See my talk from the DB Futures workshop, June 2011)

  => Get the maximum advantage of the new Oracle featutes to increase the efficiency of the SQL statements, thus reducing the load on the server (of special interest is the result set caching = zero data block reads)

  Consequently a lot more work involving close interactions between DBAs and developers is needed (in ATLAS there are more than 70 DB application responsibles)



To address the challenges ahead (cont.)

- **3) Decouple the transactional and data warehouse load**

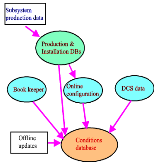
  => Use of Oracle 11g Active Data Guard as it enables read-only access to the physical standby of the primary database. The idea is that the transactional workload is served by the primary database while the heavy analytic queries and wide time range searches are routed to the standby database.

- **4) Inventory on the client applications**

  => People responsible of the major DB applications need to have the instrumentarium for analyzing the users requests on the middleware (e.g. Apache’s logs) as this info is vital for understanding how the system is used and the sources of load.

- **5) Well defined targets**

  => Chasing a moving target wastes time and at the end does not make anybody happy. There must be certain limits!



Key messages

- We need to make sure that we do not disrupt the current production services and that we introduce changes only for making the services more efficient.
- Well tuned databases are not only responsibility of the DBAs, but also of the developers. A controlled use of the resources from the end users is key factor as well.
- Special thanks to the PhyDB support for the excellent Oracle database service !