Database monitoring and service validation

Dirk Duellmann
CERN IT/PSS and 3D
http://lcg3d.cern.ch
LCG Database Deployment Plan

- After October '05 workshop a database deployment plan has been presented to LCG GDB and MB

- Two production phases
  - March - Oct '06: partial production service
    - Production service (parallel to existing testbed)
    - H/W requirements defined by experiments/projects
    - Based on Oracle 10gR2
    - Subset of LCG tier 1 sites: ASCC, CERN, BNL, CNAF, GridKA, IN2P3, RAL
  - Oct '06- onwards: full production service
    - Adjusted h/w requirements (defined at summer '06 workshop)
    - Other tier 1 sites joined in: PIC, NIKHEF, NDG, TRIUMF
LCG 3D Service Architecture

T0 - autonomous reliable service

T1 - db back bone
- all data replicated
- reliable service

T2 - local db cache
- subset data
- only local service

Oracle Streams
http cache (SQUID)
Cross DB copy & MySQL/SQLight Files

Online DB
- autonomous reliable service

R/O Access at Tier 1/2 (at least initially)
Two Main Technologies - Databases & FroNTier

- Need to provide availability information and diagnostics to different audiences
  - Database Administrators at Tier 0 and Tier 1
  - Experiment resposibles
  - Grid Deployment team

- Granularity required is very different

- Approach proposed:
  - Start from detailed (DBA level) monitoring at Tier sites
  - Extract / aggregate higher level information for experiment and grid dashboards
  - Add site test jobs (eg COOL service test) as experiment production is starting
T0 and T1 Database Monitoring

• Web based collector and user interface available as part of Oracle s/w
  - Oracle Enterprise Manager (aka Oracle Grid Control)
• A central Oracle Enterprise Manager repository at CERN has been setup to collect the status and detailed diagnostics of all 3D production clusters
  - Some sites will in parallel have the information integrated into their site local OEM setups
  - Allows to drill down to individual queries and users
  - To be used by site DBAs
• All Tier 1 sites are requested to join now
  - RAL is driving the test and documented procedure
DB Resource Usage
Lemon Integration & Weekly Reports

• For CERN Tier 0 we supply also LEMON probes
  - Aggregate DB resources usage by experiment application
  - To be used by experiment responsible
• Integrating now with new LEMON service concept
  - Color coded availability and service capacity display
• Preparing weekly resource reports to experiments and grid deployment
  - Amount of CPU & I/O used per DB application
  - Pointing out problematic DB sessions
    • Very short session, long but idle sessions
    • Very many session from a single user or host
    • Ineffective application implementations (no bind-variable, no indices)
• Used with experiment / project responsible for service sizing
STREAMS Architecture

- Capture Process
- Log Changes
- REDO Log
- Source Database
- User Changes
- Destination Queue
- Propagate Events
- APPLY Process
- Target Database (replica)
- Capture
- Propagation
- Apply
- LCRs
STREAMS monitoring

- "Home-made" scripts
  - capture, propagation and apply status
  - queues status
  - processes statistics
- STRMMON: Oracle Streams monitor tool
  - overview of the Streams activity
- Health Check report
  - information on the setup and operation of Streams
STREAMS monitoring

Display general information about each capture process

<table>
<thead>
<tr>
<th>Capture Name</th>
<th>Serial Number</th>
<th>ID</th>
<th>Number State</th>
<th>Redo Entries Scanned</th>
<th>Total LCRs Enqueued</th>
</tr>
</thead>
<tbody>
<tr>
<td>STRMADMIN_CAPTURE</td>
<td>C001</td>
<td>136</td>
<td>7 CAPTURING CHANGES</td>
<td>13394731</td>
<td>705854</td>
</tr>
</tbody>
</table>

Interval = 3, Count=1000
Logon= @ ORACLE 10.2.0.2.0
Streams Pool Size = 752M

LOG : <redo generated per sec>
NET : <client bytes per sec> <dblink bytes per sec>
Cxxx : <lcrs captured per sec> <lcrs enqueued per sec> <capture latency>
MEM : <percent of memory used> % <streams pool size>
PRxx : <messages received per sec>
Qx : <msgs spooled per sec> <msgs spilled per sec>
PSxx : <lcrs propagated per sec> <bytes propagated per sec>
Axxx : < lcrs applied per sec > <trans applied per sec> <enqueue latency>
+F : flow control in effect.
+B : potential bottleneck
+x%I x%F x%xx : <idle wait events percentage> <flow control wait events percentage> <other wait event percentage and name>
x-> : database instance name

2006-06-06 16:25:26 | d3r1-1-1 | | | MEM 6 % 752M
2006-06-06 16:25:26 | d3r1-1-1 | LOG 512 | NET 6K 0 | | | C001 0 0 3sec <0%I 0%F -> | Q46190 0 0 | PS01 0 0 <89%I 0%F -> | PS02 0 0 0 <0%I 0%F -> | MEM 6 % 752M
2006-06-06 16:25:29 | d3r1-1-1 | LOG 0 | NET 6K 0 | |F| | C001 0 0 3sec <0%I 0%F -> | Q46190 0 0 | PS01 0 0 <100%I 0%F -> | PS02 0 0 0 <0%I 0%F -> | MEM 6 % 752M

LCG 3D Status23
February 2006

Dirk Duellmann3D Meeting --- Luis Ramos
Detailed Streams Status

• Developing a detailed stream status display together with ARDA/EIS
  - Filling a gap in Oracle Enterprise Manager
• Plots for replication load (at the source DB) and individual site throughput and backlog
• Exercised as part of the replication throughput test with the Tier 1 sites
• Aim is to provide a few metrics to existing monitoring setups (eg experiment and grid dashboards)
Frontier Production configuration
FroNTier/Squid Monitoring

- SNMP based setup has been developed by FNAL
  - probes Squid cache availability and usage
  - was used during the CMS FroNTier tests
- For current LCG 3D pre-production phase this setup has been copied
  - Currently run by CMS and hosted at FNAL
  - May need to re-discuss if more experiments will require FroNTier
- Being integrated into CMS dashboard by ARDA/EIS team
Monitoring Squids W/ SNMP interface (MRTG plots shown)

Squid@CERN Requests/fetches

Squid@CIEMAT Requests/fetches

Squid@CERN In/out

Squid@CIEMAT In/out

Test: 20 Parallel CMSSW Clients @ CIEMAT
Client Side Monitoring

• LCG Persistency Framework builds on top of general database abstraction layer CORAL
  - POOL and COOL use this layer for all database access
  - Oracle, MySQL, SQLite and FroNTier are back-end plugins

• CORAL implements detailed monitoring of perceived latency of connection and query timing
  - Information would be useful to complement server side
  - Discussed interface to Mona Lisa with CMS

• So far no collection/aggregation of this information due to lack of resources
Object Diagram

Application

- IRelationaService
- IRelationaSession
- IMonitoring

DB Reporter

CSV Reporter

XML Reporter

CORAL PLUG-IN

FrontierAccess

MySQLAccess

OracleAccess

CORAL PLUG-IN

CORAL PLUG-IN
Summary

- Database services are well established at Tier 0 and being setup and tested at Tier 1s

- Database monitoring has to cover very different granularity
  - (Very) Detailed monitoring available at Tier 0 and soon all Tier 1s with Oracle Enterprise Manager
  - Working on providing relevant summaries/aggregated results for existing experiments and grid deployment
  - Some iteration on selection the right subset of metrics will be required as the production use by the experiments is ramping up

- Higher level site validation tests from grid jobs will need to be added as soon as experiment test start
  - Plan to align with the existing infrastructure on the deployment side for these tests