Status and plans on PhEDEx-LFC integration

A. Sciabà (CERN), N. De Filippis (Bari)

CMS/LCG/EGEE Task Force
Integration meeting
Aug 31, 2005
Status

- POOL LFCCatalog plugin released (2.1.0)
- Functionality tests performed
  - POOL CLI thoroughly tested
  - some bugs found, fixed in later releases (now 2.1.2)
- Performance tests performed
  - test suite provided by CMS, same used to validate any POOL catalog
  - no changes needed to use it with LFC
  - performances look reasonable, but no clear indication on whether they are up to production requirements
Test setup

<table>
<thead>
<tr>
<th>Site</th>
<th>LFC client</th>
<th>LFC server</th>
<th>Backend</th>
</tr>
</thead>
<tbody>
<tr>
<td>CERN</td>
<td>PIII 1 GHz</td>
<td>dual Xeon 2.4 GHz</td>
<td>Oracle</td>
</tr>
<tr>
<td>Bari</td>
<td>PIII 1.266 GHz</td>
<td>PIV 2.8 GHz</td>
<td>MySQL</td>
</tr>
</tbody>
</table>

- CERN is the LFC pilot installation
  - database shared among all VOs
- 10000 fake entries inserted
  - LFN-PFN pairs, mimicking typical CMS names
## Performance summary of POOL catalogs

<table>
<thead>
<tr>
<th>No. of clients</th>
<th>LFC</th>
<th></th>
<th></th>
<th></th>
<th>Oracle(*)</th>
<th>MySQL(*)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Oracle</td>
<td>MySQL</td>
<td>Query by GUID (ms)</td>
<td>Query by PFN (ms)</td>
<td>Query by GUID (ms)</td>
<td>Query by PFN (ms)</td>
</tr>
<tr>
<td>1</td>
<td>384</td>
<td>87</td>
<td>264</td>
<td>102</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>111</td>
<td>18</td>
<td>79</td>
<td>20</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>20</td>
<td>159</td>
<td>6</td>
<td>120</td>
<td>6</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>50</td>
<td>6</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>16</td>
<td>14</td>
</tr>
</tbody>
</table>

(*) Results from Jens Rehn
Remarks on LFC performances

- Substantial overhead compared to database access (performances improve a lot with parallel queries)
- Queries by GUID are very slow due to authentication being repeated for each file
- Queries by PFN as fast as with other catalogs for parallel queries
- Time to publish an XML fragment of 19 files to DB:
  - ~ 0.5 s  Oracle
  - ~ 3 s  LFC Oracle
  - ~ 2.2 s  LFC MySQL
Plans on integration with PhEDEx

- **Short-term (2-3 weeks)**
  - Configure PhEDEx with POOL-LFC interface
  - Run realistic transfers from remote sites
  - Test migration from existing POOL catalogs to LFC
  - More extensive scalability tests?

- **Medium term (1-2 months)**
  - Test LFC as a POOL file catalog for CMS analysis jobs

- **Long term (3-4 months)**
  - Stronger integration with LCG SRM SEs (DPM, dCache, CASTOR)?