Experiences with gLite RC1

Dietrich Liko IT/GD/ED

http://arda.cern.ch

Information Society

- **ARDA**
  - Prototype activities

- **How gLite has been presented to us?**

- **WMS**

- **Data management**
  - Catalog
  - gLiteIO

- **CLI**

- **Integration WMS - Datamanagement**

- **Conclusions**

- **ARDA develops prototypes with the experiments**
  - ALICE: Interactive Analysis using PROOF
  - ATLAS: High Level Service DIAL
  - CMS: Physh
  - LHCb: GANGA

- **ARDA provides feedback to the middleware developers**
  - Early access to developer prototype
  - Feedback to developers
  - Testing of gLite components

- **I have tried to collect a lot of material –maybe I did not always understand all aspects**

- **The middleware is rapidly moving and it is hard to keep track**

- **All achievements in the following pages are by the middleware developers and the testers**

- **All errors and misunderstandings are due to my fault**

- **Prototype test bed**
  - Run by the developers in JRA1
  - Not all features immediately useable for us
    - Documentation, communication etc.

- **Early Start in May**
  - First components (including Alien)

- **WMS (October)**

- **Data management (November)**
  - Catalog
  - gLiteIO
  - CLI

- **End January gLite RC1 was defined**

- **We are seeing**
  - active development
  - ongoing integration of components

- **Excellent collaboration with developers**

- **Noticeable improvements on weekly basis**
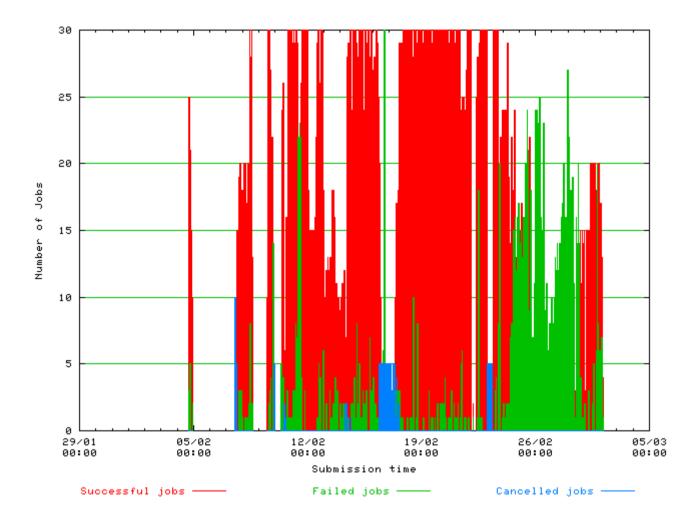
- **Delivered in October**
  - Internal reorganizations (CondorC)
  - "Pull" mode (CEMon)
  - To the user similar to LCG WMS

- **New features**
  - DAGs
  - Integration with VOMS
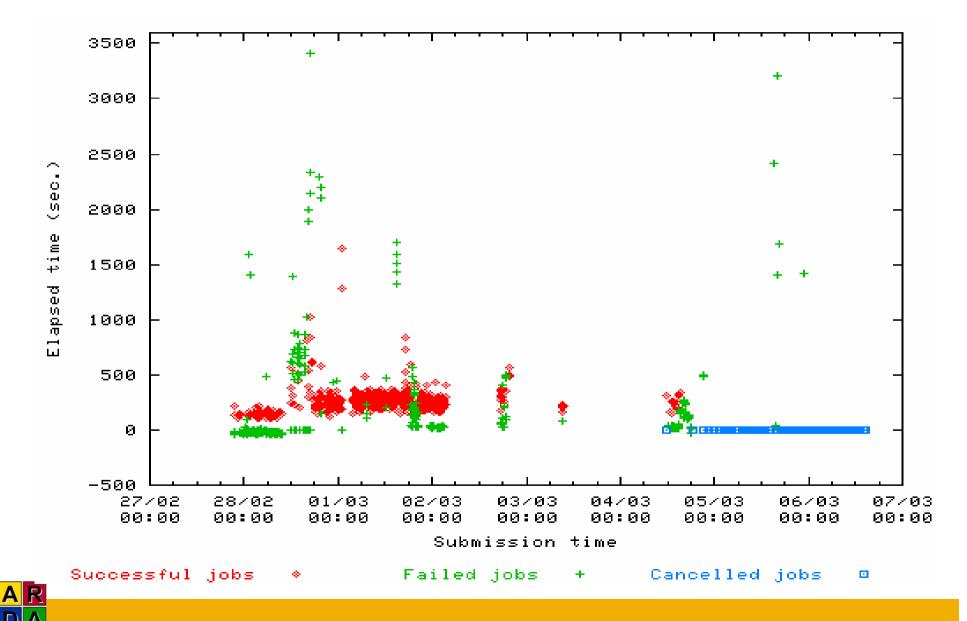 - Integration with StorageIndex

- **Web service interface not for RC1**

- **Work in progress**
- **Integration effort is ongoing**

- **Some problems are present**
  - A number of jobs abort (~10%) – Condor problem
  - PBS CE did not seem to work

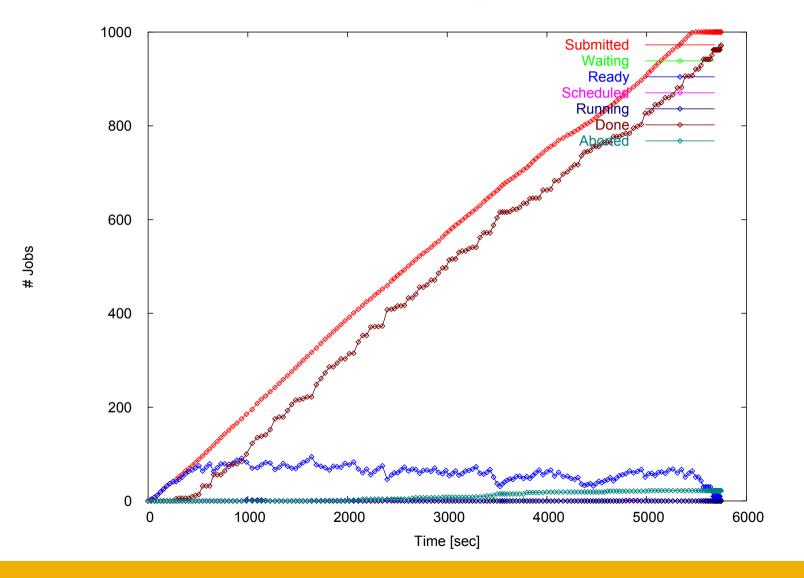- **A jobs needs usually ~300 seconds to be executed**

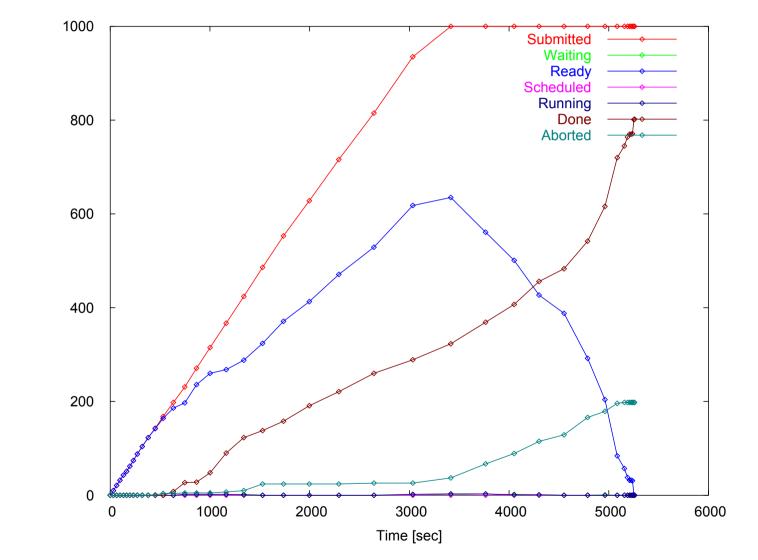WMS – Last Weekend

ARDA WMS profile -

- **Same problems as before**

- **No obvious problems due to file placement**

- **Submission speed**
  - Single thread: 5 secs/job
  - 10 threads: 0.5 to 2 secs/job observed
  - RB usually to keeps up with the speed

ARDA WMS profile -

Legend:
- Submitted
- Waiting
- Ready
- Scheduled
- Running
- Done
- Aborted

Y-axis: # Jobs (0 to 1000)

X-axis: Time [sec] (0 to 6000)

- **In general the WMS performs quite reliable**

- **Not always clear if all observed aspects are problems – this is a developer test bed and we do not always know what is being done.**

- **There are some problems that should be solved before going the RC1**

- **gLiteIO**
  - POSIX like IO package
  - Integrated with catalog

- **Fireman catalog**
  - Oracle & MySQL
  - Full authorization model
  - ACL

- **RFT/FPS**
  - Not tested by ARDA at this point

- **Simplifies the grid world for the user**
  - Catalogs, SRM
  - IO protocols: rfio, dcap, ….

- **Uses a client - server model**
  - Security

- **Scalability and reliability has to be demonstrated**
  - Consistency catalog and actual data on disk
  - High load
  - Interactive usage (many users accessing many files using ROOT)

- **Tests by Andrei Demichev**

- *Intention:* To put/get ~1000 **SMALL** files of random size in the range **0 10 Kb** in maximally **gentle** regime, i.e., only **one** file was put/get during the cycle.

- End 04
  - Average number of put/get cycles done in a row before an error appeared: 64
  - Max. number of put/getcycles done in a row before an error appeared: 256

- **Situation improved …**

- Tests are now run by JRA1

- **What load the server can stand ?**

- **How many users can one server support ?**

- **Where do we want to abstract the IO protocol ?**
  - Application
  - Client Library (LCG GFAL)
  - Intermediate server (gLiteIO)

- **Integration with ROOT, POOL etc**

- **Includes full security model**
  - Users, groups, permissions, ACL

- **Oracle & MySQL implementations available**

- **StorageIndex interface to WMS**

- **For RC1: Global catalog**

- **In the future: Local catalog**

- **Catalog**
  - Started last year with Victor Pose
  - Now Craig Munro

- **Work in progress**
  - New features are enabled weekly
  - Authentication/Authorization
  - Not always stable service

- **We are reporting all problems to Savannah**
  - Concurrency problems
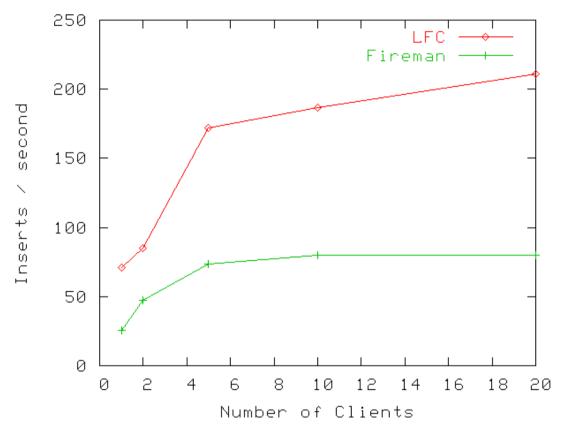 - Functional problems

- **The following plots are not final results, but they represent a glimpse on work in progress.**

- **The setup of tests shown here cannot be compared**

  - All of the tests are based upon the original LFC test suite and compared with the original LFC results. Fireman is running on a box shared with other services with 5 threads in Tomcat compared to 20 for LFC with it's own box.

- **The developers are working hard to improve their components**

- **We are working hard to improve our tests and our understanding of the software**

- **We think it is nevertheless important to present the state to the experiments**

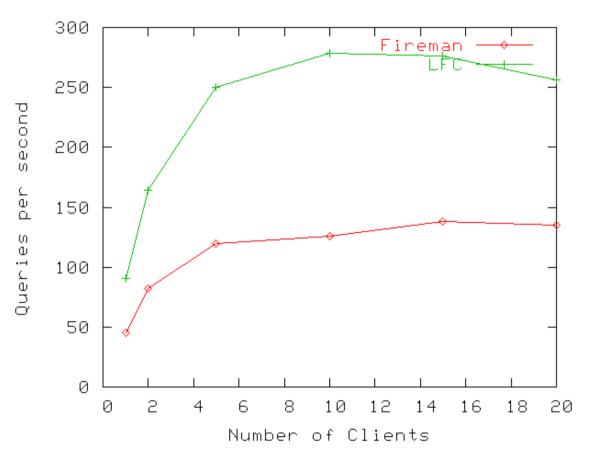- **We are sure they will trust our common effort more if we show the current status**

- **Inserts 1000 entries using multiple clients and measure the rate of insert**
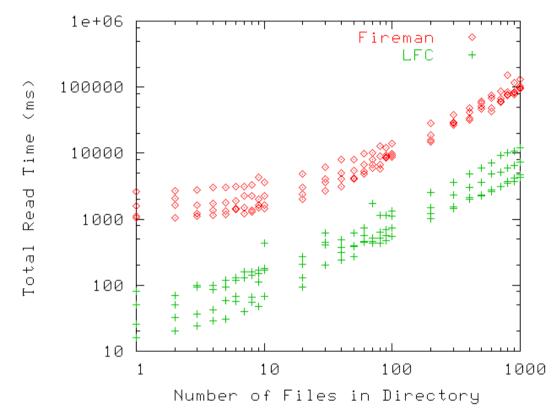- **MySQL about 10/entries**

- **Query test repeatedly looks for a random entry in a directory containing 1,000,000 entries and measures the time to find the entry.**

- **Read performance measures the time the readDir operation takes for increasing number of files in the directory and threads. Results are shown for 1,2,5 and 10 threads with Total read time decreasing for increasing threads.**
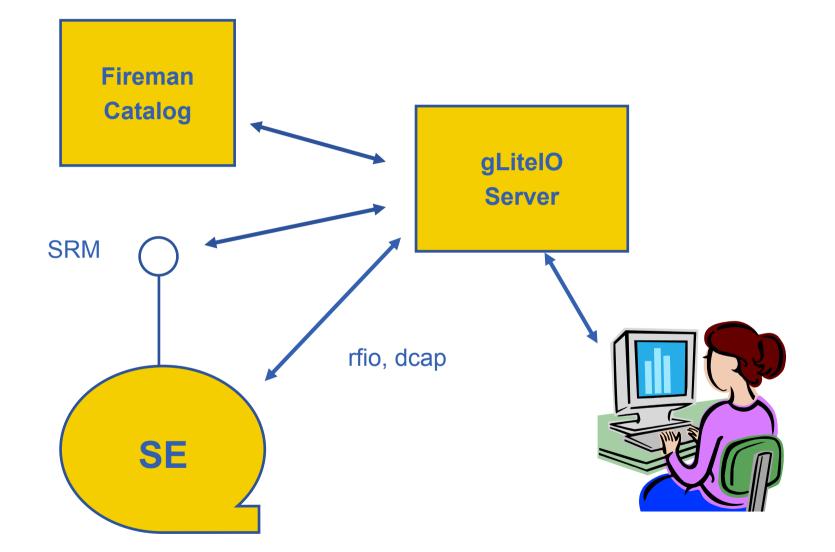
- **The following slides represent my personal understanding of concepts**
  - Please correct me if I misunderstood

- **Why gLiteIO ?**

- **Why the StorageIndex ?**

- **What are the plans of the experiments**

- **Some personal conclusions**

Fireman
Catalog

gLiteIO
Server

SRM

SE

rfio, dcap

- **Plus**
  - Catalog has full security model
    - acl, owner, groups
  - Allows to keep track of user files on the grid

- **Minus**
  - System with many components
  - Has to be very reliable
  - Performance ?

- **Other aspects will have to be addressed**
  - Quota management
  - Consistency catalog/SE

- **StorageIndex**
  - Compact central catalog that scales better then the full catalog

- **For RC1**
  - Global File Catalog
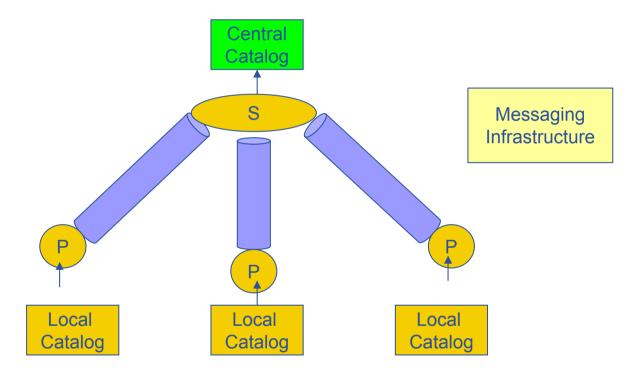 - StorageIndex alternative interface to the catalog for the WMS

- **In the future**
  - Local fireman catalogs
  - Central storageIndex
  - Messaging Protocol to keep catalogs in sync

- **From the last ARDA workshop**

- **Not part of RC1**
- **Nevertheless it should be demonstrated at some point to the community**

My understanding of various discussions …

- **Data will be distributed on dataset basis**
  - Seems natural
  - Discussion on the size

- **Only Central Dataset repository necessary**
  - Experiment specific ?

- **For experiment data acl's might not be necessary**
  - Reliability
  - Performance
  - GUID lookup necessary

- **Secure catalog allows individual grid users to store files on the grid**
  - Very important in the long term

- **There are many aspects to be demonstrated**
  - Performance and reliability of an SE with gLiteIO
  - StorageIndex concept

- **Experiment plans are as well important**
  - LHC is coming
  - Should be taken into account in some form on the LCG preproduction test bed

- **Observe the development**
  - XROOTD
  - Will we have at some point integrated SE that understand grid users

- **gLiteIO**
  - Interesting features
  - To be a success it has to be rock solid also under high load

- **Catalog**
  - Basic features have been demonstrated
  - Not always stable service in the last weeks
  - Security and access control has still to be demonstrated
  - Performance comparisons with LFC or GlobusRC have to be assessed in a controlled way

- **Since two weeks available**

- **Clear improvement, fundamental for end user**

- **Functional equivalence to LCG Utilities**
  - Not all functionality yet available

- **We hope that the UI development will be continued in this spirit**

- **Still a lot of work ahead to shrink the size**
  - LCG: 445 MB - gLite (today) 250 MB - Nordugrid: 14MB

- **Exercise**
  - Register a file using gLiteIO
  - Place a job on the grid driven by the location of this file
  - Read the file by the job on the worker node using gLiteIO

- **First successful exercise by us two weeks ago**
  - Using the MySQL catalog

- **Most difficulties reported by us were addressed in short time**
  - WN setup
  - Inconsistent notation
  - Did everybody know ?   **lfn:///myfile.dat**

- **Other problems seem to be more subtle**
  - StorageIndex using Oracle

- **Prestaging of data using RFT**

- **Integration with a package manager**

- **Why not some Alien like features like lfn in the sandbox etc.**

- **gLite RC1 is taking shape**

- **A lot of progress can be observed in the last weeks**

- **Still some way to go**

- **There is always the wish to add further functionality, but stability and reliability are the fundamental basis for the product to be accepted by the community**