Experiences with Testbed1, plans and objectives for Testbed 2

Testbed retreat 27-28th August 2002

Julian.Linford@esa.int

EU-DataGrid Project
Work Package 9 – EO Applications

Experience with Testbed v1.2

- User Interface installed RH 6.2 & 7.2
- Basic job submission tests (single short jobs using input sandbox and CERN UI)
- Data replication tests ongoing
  - problems with GDMP_CONFIG_FILE
- Started installing 1.2 SE
- Intensive use of 1.2 Testbed planned starting September
  - carry out processing of 1 month of GOME data (about 350 15Mb files 5.25Gb), data needed by IPSL for Validation tests
  - upgrade CE to 1.2 + interface with AFS/LSF
  - complete installation of 1.2 SE + interfaces with the ESA MSS system
- Work will concentrate on preparing several high-profile demonstrations
  - implementation of GOME Data processing and validation Use Case
  - EO WebMap Portal
    - visualisation of global Ozone measurements
    - on-demand product processing using EDG services

Experience with Testbed v1.1

- 1.1.4 Middleware Installed at ESRIN
  - User Interface
  - Network Monitoring Tools
  - Computing Element (with PBS and LSF)
- Has been used to partially carry out the basic GOME use case:
  1. Transfer Level1 (raw) data to the Grid Storage Element
  2. Register Level1 data with the Replica Manager
  3. Submit jobs to process the Level1 data, produce Level2 data products
     - Jobs running on the CEs locate Leve1 data by using the **BrokerInfoAPI**
  4. Repeat step 1-3 for level 2 products
     1. Transfer Level2 data products to the Storage Element
     2. Register Level2 data products with the Replica Manager
     3. Submit jobs to the Grid to validate Level2 data products
  5. Retrieve validation results and visualize at the User Interface

Commands we used

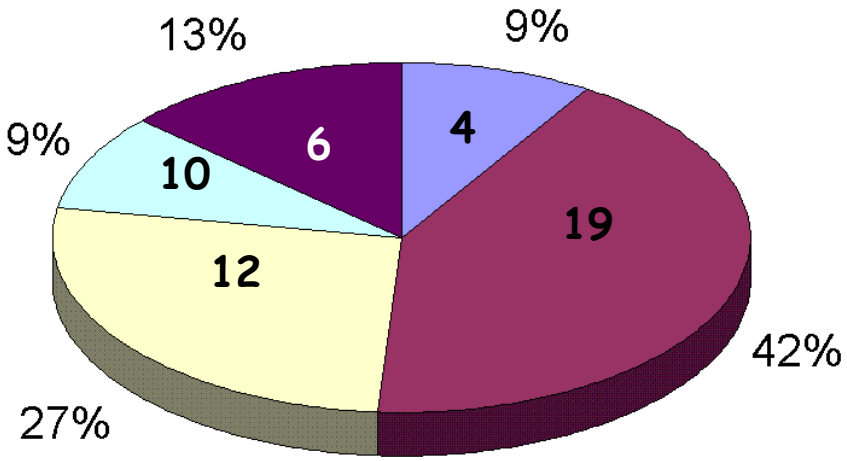
- Job execution
  - ✓ dg-job-list-match
  - ✓ dg-job-submit
  - ✓ dg-job-status
  - ✓ dg-job-get-output
- Data management
  - ✓ gdmp_register_localfile
  - ✓ gdmp_publish_catalog
  - ✓ gdmp_get_catalog
  - ✓ gdmp_replicate_get

Results

- Application Environment installation not straightforward
  - need to contact sites directly to verify / fix problems
- Perceived general instability of the testbed
  - high incidence of unrecoverable errors, intermittent errors
- Job submission comands, execution cycle basically OK
  - but need better support for handling multiple simultaneous jobs
  - not easy for apps to work with CLI
- Data management commands not easy to use
  - complex replication sequence
  - unreliable, intermittent working
  - difficult to diagnose cause of malfunctions

Meeting EO Requirements

- EO Requirements have been surveyed and matched against the testbed functionality (D9.6 Scaling Study activity)
- Analysis of 45 basic requirements (several were grouped together)
  - 4 Satisfied
  - 19 Partially satisfied
  - 12 Expected in future releases
  - 4 Planned in future releases
  - 6 Need to be verified



- Overall basic functionality of job submission and data replication considered satisfied in TB1
- Although the commands work, the testbed has not yet reached the required production quality

EO requirements priorities

- Reliability
  - Improve job failure rate under high load conditions
  - Increase robustness and fault tolerance
  - Improve configuration / installation
  - Remove / avoid single points of failure
- Documentation
  - Needs constant revision to keep up with software changes
  - Installation manual, user manual should be mandatory in delivered RPMs
  - Verified and approved by quality control / testing measures

EO requirements priorities

- Usability
  - New Data management s/w not yet very well understood
    - Overlapping command sets
      - GDMP
      - Replica Manager
      - Replica Catalog
      - Interface to MSS
    - Need for clear procedures / instructions
  - Cryptic error messages
  - Need better error recovery (core dumps are a show-stopper)
  - Need built-in error handling and fault tolerance

EO requirements priorities

- Application interfaces
  - low-level commands require application functional layer

    and ∴ need to be designed appropriately
  - middleware command interfaces subject to change
  - apps will need some minimum backward compatibility
- Site uniformity
  - a job should produce the same result regardless of where it executes
  - it should not require hard coded values for a specific site
  - avoid end-users having to contact sites directly

EO requirements priorities

- Automatic job decomposition based on input dataset
  - need use case / examples
- Brokerinfo use cases
  - method to locate the replicated files locally on the CE
- SE storage (for data & scratch space) management
  - query amount of space available
  - advance reservation
- Create / destroy VOs
  - + groups within VOs
- Integration of EO archives and catalogue systems

EO Use Case File Numbers

1 Year of Gome data

Data	Number of files to be stored and replicated	Size
Level 1	4,724	15 Mb
Level 2 NNO (ESA)	9,448,000	10 kb
Level 2 Opera (KNMI)	9,448,000	12 kb
Validation Lidar (IPSL)	12	2.5 Mb
Total:	**18,900,736**	**267 Gbyte**

Gome has a data set of 5 years

Gome is relatively small (in both size and number of files)

EO requirements for Testbed2

- SE storage management policies
  - need for standard – even automatic - procedures for freeing space on the SE
- Capability to store user-defined application metadata in RC
  - use metadata keys as alternative to to LFN to describe input data
- RB support for data pipelining
  - Ability to specify input data which will be produced as a result of a previous processing step
- Retrieval of QoS measures for data access, storage and processing (lost data at RAL!)
- Per VO / user quotas on job submission, RC & SE usage

Quality Objectives

- Should be carefully selected to make a rapid impact on production Testbed stability and reliability
- Minimal disruption to the production testbed during upgrades, patches & site / service outages or re-configurations
- Reduce priority of new functionality until existing infrastructure is stable and proven
  - priority to bug fixes and basic system enhancements
- Develop & apply documentation suite & standards
- Acceptance test procedures

Quality Objectives

- QA representatives should actively promote use of fault tolerance, defensive programming techniques, diagnostic facilities, etc.
  - sw development cycle should include testing and validation plans for each unit
  - make reliability a major design objective
- Quality control checklist for single RPMs to include
  - testing and verification details log
  - comprehensive installation and user manual
  - automatic installation and configuration scripts
  - test and verification scripts

A few suggestions

- Documentation suite and standards
  - dedicated document writers
- Reference test suite
- Acceptance tests – automatic procedures
- Towards automatic monitoring & anomaly detection
  - background information gathering by test probe jobs
- Testbed status / news / info update
  - e.g. like the login banner which reports current status of sites & services down, critical bugs, etc.
- Clear instructions on when to test and what to test