Table of Contents

LHCb DAST Instructions

Introduction

This page provides info for the LHCb Distributed Analysis Support Team (LHCbDAST). Ulrik Egede from the Ganga team organize this effort. DAST is the first point of contact for all distributed analysis questions.

- Help is provided through the Distributed Analysis mailing list Distributed Analysis mailing list ⧉
    - ♦ The shifter is subscribed to this list and will see your query. Please do not mail the shifter directly.
    - ♦ Messages coming into the forum will be responded to within a few hours during the shifters institutional daytime Moday to Friday. Outside these times support is voluntary,

The Team

- Cedric Potterat
- Jack Wimberley
- Jason Andrews
- Jibo He
- Mark Slater
- Michael Alexander
- Patrick Owen
- Robert Currie
- Ulrik Egede
- Carlos Vázquez Sierra

Shift calendar

Shifts are booked through the LHCb shifters database ⧉

Previous Shift Data

- Updates from the Ganga Devs Out-of-date Last update: 11/05/2015
    - ♦ Some users observe a non-starting Ganga 6.1.13. This appears to be fixed by running: 'find ~/gangadir/repository/${USER}/LocalXML/6.0/jobs/* -empty -type d -delete'
- DAST shift report from the DAST shifter Out-of-date Last update: 04/03/2016
    - ♦ Most issues were related to Ganga. A few issues with warning messages, persistency of jobs if there's a crash and possible lock ups were reported. These have (hopefully) now all been dealt with and should be available in 6.1.17. LHCbTasks still has issues but these are going to be addressed this week. I suspect the new release of Ganga will be in the Dev area first but an email will be sent out to give info. Other than that there was a Dirac service issue due to a restart of at least on of the voboxes and a problem with a corrupted userkey.pem file.

References for shifters and users

General references

- LHCbSoftwareTutorials
- FAQs
- LHCbComputing
- LHCb Q&A ⧉

Ganga References

- Ganga User Guides🗗
- Distributed Analysis mailing list🗗
- Ganga issues on Github🗗
- Ganga on LHCb Q&A🗗
- Ganga channel on LHCb mattermost🗗

Users Section

What do we need to know

Contact the shifters via the Distributed Analysis mailing list🗗

When reporting problems regarding distributed analysis, **please include all of the following information** so that the shifter can work most effectively to solve the problem. If the report is incomplete, *it will take us that much longer to help you with the problem.*

- **Summary:** Provide a descriptive summary of the issue.
- **Environment** Describe Ganga version used, if you run at CERN or on your laptop etc.
- **Steps to Reproduce:** Detail the exact steps taken to produce the problem. Where applicable, provide links to jobs and outputs, as well as jobOptions and the command lines you used to submit the jobs.
- **tracking information** Provide the Dirac ID of some representative problematic jobs.
- **Expected and Actual Results:** Describe what you expected to happen when you executed the steps above, and what actually occurred.
- **Regression:** Describe circumstances where the problem occurs or does not occur, such as software versions or specific sites.
- **Notes:** Provide additional information, such as references to related problems, workarounds and relevant attachments.

In case of any problems related to Ganga e.g. a crash with a traceback or a mysteriously failed job we would recommend using the `report()` tool to help the operations team to help you. The output can be viewed via the gangamon.cern.ch🗗 page. More information about this is available in the FAQ.

Shifters Section

Shifter Daily Procedure

Check the downtime calendar to be familiar with what issues could come up because of this (sites not available etc). Downtime calendar🗗

Shifters should make an effort to check their emails about once every hour and should strive to reply to mails from the lhcb-distributed-analysis list, and where appropriate forward the email to the appropriate experts.

When a user doesn't provide a lot of information on a problem and more information would be needed to work out what's going wrong please try and prompt them to read:
https://twiki.cern.ch/twiki/bin/view/LHCb/FAQ/GangaLHCbFAQ#Procedure_for_reporting_errors

The summary of each shift (per-week) should be filled in as described in the section on: Previous Shift Data. This should normally be a few lines summarising the most important information for the following shifter to be aware of.

e.g. There is a problem with MyApplication v4r5 on the grid users should revert to v4r3 until a fix is ready

The ganga devs will attempt to put any common work arounds here which will be useful for the shifter to know to reply to user queries

Computing Operations Meetings

It is advised that the DAST shifter should where possible try to attend the Computing Operations meetings preferably on Mondays and Fridays, or at least read the minutes to be aware of any issues which could affect user jobs on the grid.

- Monday, Wednesday and Friday Computing Operations meetings⊡.

Escalations

Issue	Escalate to	Comment
Ganga bugs	Ganga issues on Github⊡	or contact lhcb-ganga@cernNOSPAMPLEASE.ch
DIRAC bugs	LHCBDIRAC JIRA⊡	or contact lhcb-geoc@cernNOSPAMPLEASE.ch
General Grid problems	The best person to escalate issues regarding grid stability is to escallate this to the GEOC	contact lhcb-grid-geoc-oncall@cernNOSPAMPLEASE.ch or lhcb-geoc@cernNOSPAMPLEASE.ch
Login and environment issues	LHCb scripts JIRA⊡	
LHCb BookKeeping		contact lhcb-bookkeeping@cernNOSPAMPLEASE.ch
Data management		contact lhcb-datamanagement@cernNOSPAMPLEASE.ch
Tier 1 issues	Problems may be reported dirtectly to the email lhcb-_tier1_-contact@cernNOSPAMPLEASE.ch (replacing *tier1* with the site) if the problem is clearly site specific	
Site or grid issues in general	Phone CERN ext 77714 between 8am and 10pm or email lhcb-grid@cernNOSPAMPLEASE.ch	This is the production shifter
CMT	Marco Clemencic or Hubert Degaudenzi	
Core software related issues	Contact Marco Cattaneo	

Permissions and accounts required

To be able to deal with the shifts in an effective way, you need to have certain permissions and accounts:

- Grid certificate that is valid and in the LHCb VO
- Permission to look at other peoples jobs in the DIRAC Monitoring. (this should be possible through the lhcb_shifter role)
- An account for the Production e-Log⊡. Just go there and create one.
- An account in JIRA to allow you to file and monitor bugs.
- Be signed up to the lhcb-dast@cernNOSPAMPLEASE.ch mailing list

Checking DIRAC services.

You can open a Ganga instance and then issue the commands

```
In [1]:Dirac().debug()
```

and you should get an output like below. Any non-OK messages apart from the "normal" errors below might indicate a problem.

```
DataManagement/StorageElementProxy: URL for service DataManagement/StorageElementProxy not found
ProductionManagement/ProductionRequest: OK.
DataManagement/DataLogging: URL for service DataManagement/DataLogging not found
Framework/SystemLogging: OK.
DataManagement/FTSManager: OK.
Bookkeeping/BookkeepingManager: OK.
ResourceStatus/Publisher: OK.
DataManagement/FileCatalog: OK.
Framework/ComponentMonitoring: OK.
Framework/Notification: OK.
Framework/Monitoring: OK.
WorkloadManagement/JobStateUpdate-9137: URL for service WorkloadManagement/JobStateUpdate-9137 no
ResourceStatus/ResourceStatus: OK.
RequestManagement/ReqManager: OK.
WorkloadManagement/OptimizationMind: OK.
Configuration/Configuration: URL for service Configuration/Configuration not found
WorkloadManagement/JobStateUpdate: OK.
Framework/Plotting: OK.
Transformation/TransformationManager: OK.
RequestManagement/ReqProxy: URL for service RequestManagement/ReqProxy not found
Framework/BundleDelivery: OK.
Accounting/DataStore: OK.
WorkloadManagement/Matcher: OK.
WorkloadManagement/WMSAdministrator: OK.
DataManagement/TransferDBMonitoring: URL for service DataManagement/TransferDBMonitoring not foun
DataManagement/FileCatalog-OLD: URL for service DataManagement/FileCatalog-OLD not found
Framework/SecurityLogging: OK.
StorageManagement/StorageManager: OK.
DataManagement/StorageElement-1: URL for service DataManagement/StorageElement-1 not found
Framework/SystemLoggingReport: OK.
ResourceStatus/ResourceManagement: OK.
RequestManagement/RequestManager: URL for service RequestManagement/RequestManager not found
Configuration/Server: OK.
WorkloadManagement/JobManager: OK.
WorkloadManagement/SandboxStore: OK.
DataManagement/StorageElement: URL for service DataManagement/StorageElement not found
Accounting/ReportGenerator: OK.
Framework/UserProfileManager: OK.
DataManagement/DataIntegrity: OK.
DataManagement/DataUsage: OK.
DataManagement/StorageUsage: OK.
WorkloadManagement/JobMonitoring: OK.
Framework/Gateway: URL for service Framework/Gateway not found
WorkloadManagement/JobStateSync: OK.
Framework/SystemAdministrator: OK.
DataManagement/RAWIntegrity: OK.
Framework/ProxyManager: OK.
RequestManagement/RequestProxy: URL for service RequestManagement/RequestProxy not found
```

Potential causes of performance issues

From Rob:

There has been a lot of discussion about the performance problems that some users are seeing in Ganga.

I'd like to shed some light on a few possible things which can lead to this:

1) A system under high load. There's nothing we can do to speed up running on a system with slow afs access. I would encourage users to try using Ganga on a system other than lxplus if they suspect they're experiencing this.

2) Users running over several thousand LFNs. Users running over many thousands of LFN in a single job may experience ganga slowing down dramatically. This is an unfortunate consequence of the move to inputdata and is something that is on the immediate radar for the Ganga devs but cmake support is currently taking priority. Running jobs which don't require extremely large numbers of LFNs are actually faster on the latest versions of ganga.

If possible users should consider restricting the number of LFN per job if they want to speed up the responsiveness of the latest ganga. (I'd say limit it to between 500-1000 LFN per job if possible as this should be much quicker)

There is an open task relating to this: https://github.com/ganga-devs/ganga/issues/68

3) The Ganga monitoring loop being slow/stalling. Unfortunately the monitoring loop has to load a job fully into memory before it can run over it. This process is significantly slowed down due to #2 above. The monitoring loop in Ganga has been observed to stop registering new jobs during interactive sessions. Unfortunately the best advice is to wait and check on the queues system to see if Ganga is busy behind the scenes. We're hoping to work on improving the monitoring system over the next few months.

If Ganga has stopped updating jobs and doesn't appear to be busy with items in the queues, wait 5 min and restart Ganga if nothing happens.

4) Issues due to grid weather There's nothing we can do to avoid this from Ganga's side.

-- UlrikEgede - 12-Oct-2015

This topic: LHCb > LHCbDAST
Topic revision: r36 - 2017-06-19 - CarlosVazquezSierra