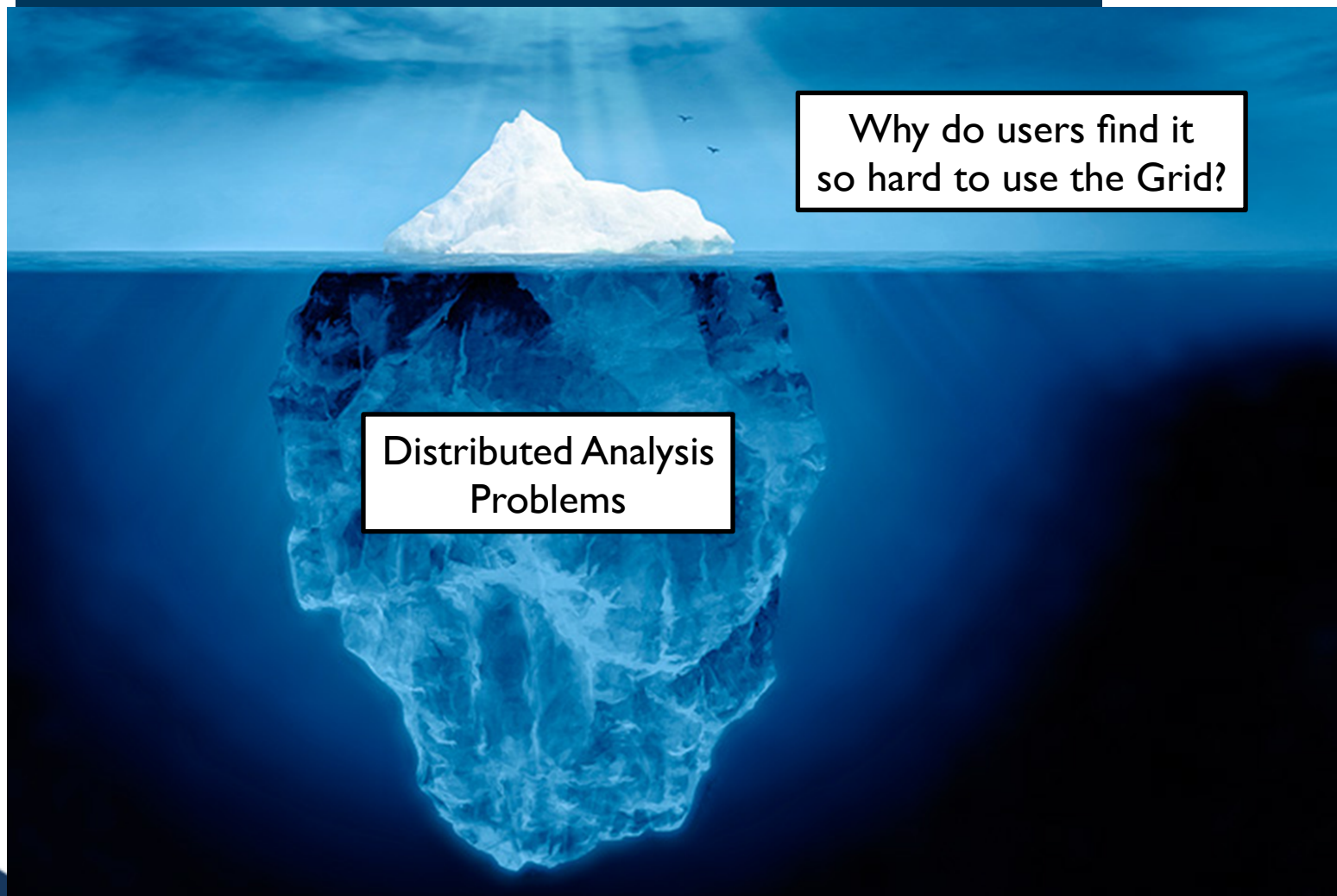


# Distributed Analysis Report

Alastair Dewhurst, Federica Legger



# Distributed Analysis



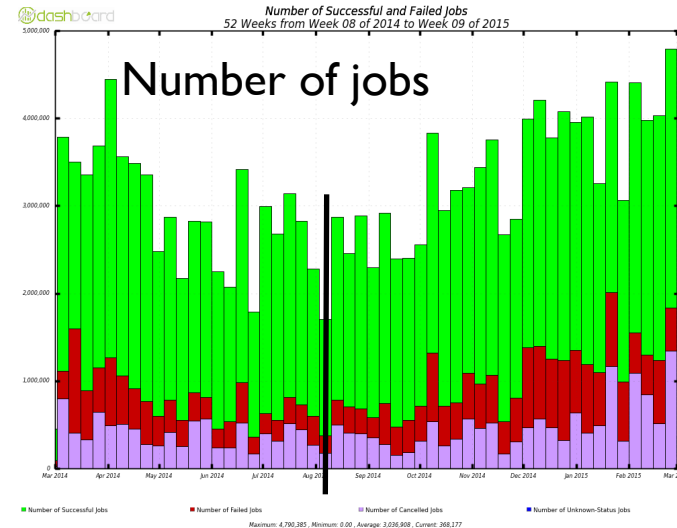
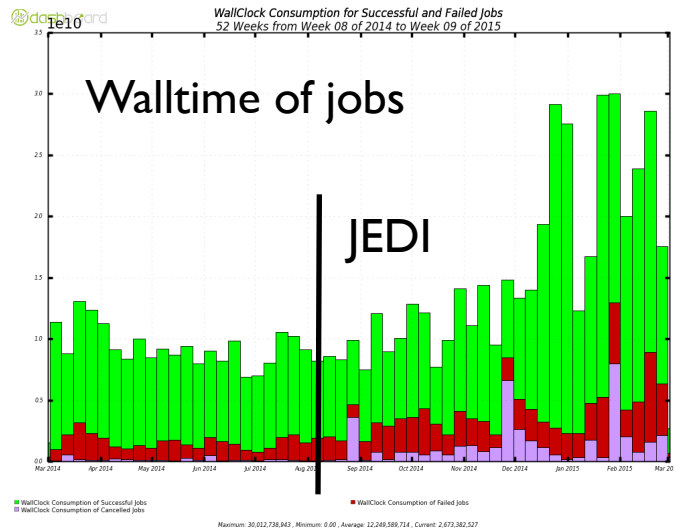
Why do users find it  
so hard to use the Grid?

Distributed Analysis  
Problems



# Cancelled jobs

- Cancelled jobs waste a significant fraction of wall time.
- Significant increase since introduction of JEDI
- Caused by users cancelling jobs.
- **Long user tasks are getting stuck!**



# Cancelled jobs(2)

<http://bigpanda.cern.ch/task/4854180/>

| Task 4854180: user.dbertsch.12All.muons.02/ |        |       |              |                      |             |                                 |                  |             |       |          |        |
|---------------------------------------------|--------|-------|--------------|----------------------|-------------|---------------------------------|------------------|-------------|-------|----------|--------|
| Task ID                                     | Jobset | Type  | WorkingGroup | User                 | Task status | Ninputfiles   finished   failed | Created          | Modified    | Cores | Priority | Parent |
| 4854180                                     | 10578  | analy |              | David Edwin Bertsche | running     | 31962   27591 (86%)   3191 (9%) | 2015-02-18 09:29 | 03-16 22:52 | 1     | 1000     |        |

| States of jobs in this task (merge jobs excluded) <a href="#">Show all jobs</a> |         |         |          |           |           |      |          |         |         |              |          |        |           |         |
|---------------------------------------------------------------------------------|---------|---------|----------|-----------|-----------|------|----------|---------|---------|--------------|----------|--------|-----------|---------|
| defined                                                                         | waiting | pending | assigned | throttled | activated | sent | starting | running | holding | transferring | finished | failed | cancelled | merging |
| 1                                                                               |         |         | 22       |           | 14        |      |          | 104     | 4       |              | 13191    | 35894  | 435       |         |

- Example: User asked why is my job taking so long. Should I cancel and resubmit it?
  - Task has been running for a month even though it has reasonably high priority.
  - A significant fraction of failures caused by obvious site problem(s). Possibly confuses JEDI?!

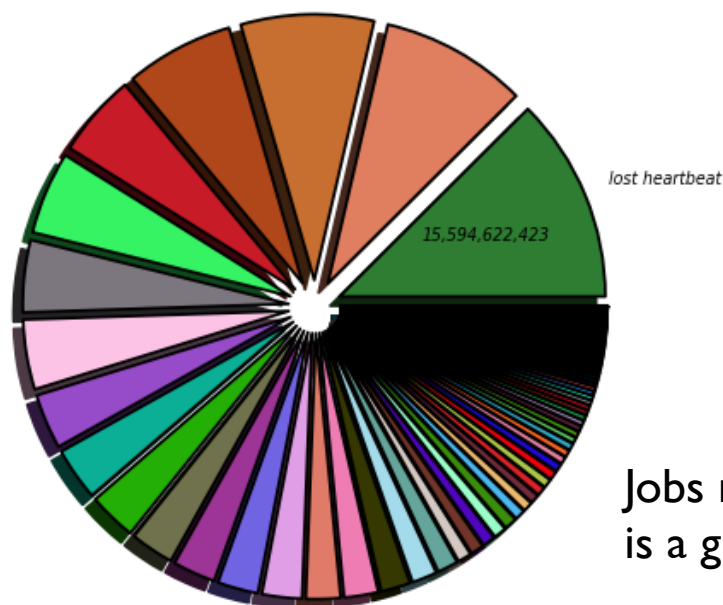


# Failed jobs (walltime)



WallClock Consumption of Panda Failed jobs by ExitCode (Sum: 125,086,658,098)

Lots of failures because of batch limit exceeded



Get / Put errors still major source of failures

Jobs running out of memory is a growing problem

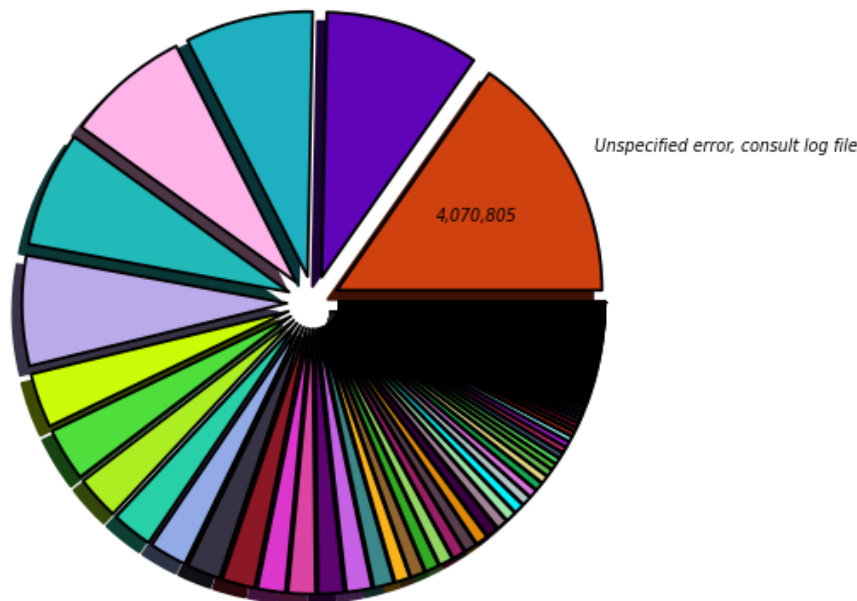
|                                                                                                                                 |                                                                                                                                          |
|---------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------|
| lost heartbeat (15,594,622,423)                                                                                                 | Get error: Staging input file failed (11,145,468,511)                                                                                    |
| Athena crash - consult log file__Reached batch system time limit (10,116,308,835)                                               | Reached batch system time limit (8,228,980,843)                                                                                          |
| Put error: Error in copying the file from job workdir to localSE (6,322,589,691)                                                | Athena crash - consult log file (6,293,342,944)                                                                                          |
| Athena crash - consult log file__Athena ran out of memory (5,484,113,351)                                                       | No events to be processed in the input file - skipEvents is higher than actual event number__Looping job killed by pilot (4,188,040,232) |
| Not documented. Exitcode: 15__Looping job killed by pilot (4,188,040,232)                                                       | Athena crash - consult log file__cancelled (4,017,498,764)                                                                               |
| Athena crash - consult log file__Looping job killed by pilot (3,665,999,067)                                                    | Unspecified error, consult log file (3,322,104,940)                                                                                      |
| Transform received signal SIGTERM__Looping job killed by pilot (3,248,170,898)                                                  | Looping job killed by pilot (3,097,017,203)                                                                                              |
| Job killed from the batch system: SIGTERM (3,033,190,935)                                                                       | Put error: Local output file missing (2,620,546,990)                                                                                     |
| New trf: Transform received signal SIGABRT; Old trf: Athena core dump or timeout, or conddb DB connection error (2,598,394,000) | Athena crash - consult log file__Reached batch system time limit (1,905,705,707)                                                         |
| Transform received signal SIGTERM__Reached batch system time limit (1,905,705,707)                                              | plus 745 more                                                                                                                            |



# Failed jobs (number)



Panda Failures by ExitCode (Sum: 26,670,467)



- Unspecified error, consult log file (4,070,805)
- Get error: Staging input file failed (2,077,962)
- Expired three days after submission (1,884,043)
- DQ2 server error (885,944)
- lost heartbeat (712,874)
- Get error: Failed to get LFC replica (607,724)
- Transformation not installed in CE (532,872)
- TRF\_SEGVIO - Segmentation violation (440,900)
- Payload core dump (410,872)
- Reached batch system time limit (241,185)

- Athena crash - consult log file (2,515,526)
- Put error: Local output file missing (2,014,783)
- Get error: Replica not found (1,787,991)
- Put error: Error in copying the file from job workdir to localSE (863,260)
- Athena crash - consult log file: Athena ran out of memory (665,740)
- New trf: Transform received signal SIGABRT; Old trf: Athena core dump or timeout, or conddb DB connect exception (542,76)
- Undocumented TaskBuffer Error Code : taskbuffer113 (448,111)
- Athena core dump or timeout, or conddb DB connect exception (416,116)
- New trf: Transform received signal SIGABRT; Old trf: Athena core dump or timeout, or conddb DB connect exception (416,116)



# DDM server errors

---

- There were 885 000 DDM server errors in March!!
- Vast majority of errors caused by incorrect characters in dataset names
  - Users were trying wildcards (\*) and colons (scopes) when they weren't needed.
- Tadashi added a check when the task is submitted:
  - <https://its.cern.ch/jira/browse/ATLASPANDA-171>
- Documentation was not clear enough.
- Small number of HC jobs with errors that should be investigated.



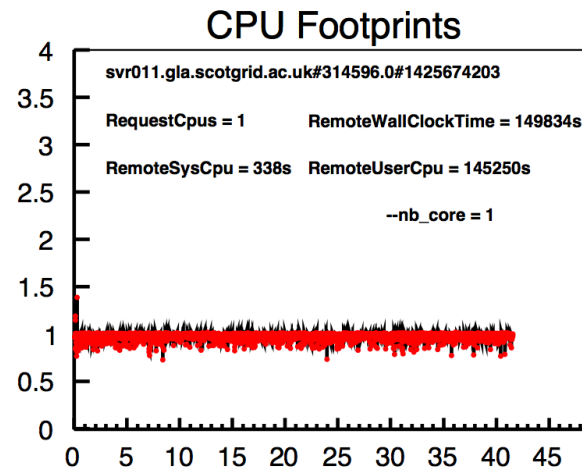
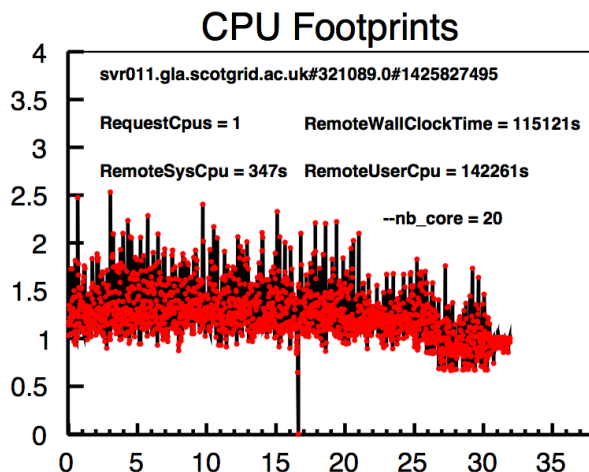
# Documentation

- We have two Panda wikis:
  - <https://twiki.cern.ch/twiki/bin/view/PanDA/PandaAthena>
  - <https://twiki.cern.ch/twiki/bin/viewauth/AtlasComputing/PandaAthena>
- Tadashi is updating the 'PanDA' twiki but the 'AtlasComputing' twiki is linked from most other ATLAS twiki pages.
- Need to consolidate twiki pages.
  - Do NOT repeat mistakes for Rucio.



# MadGraph (Event)

- The MadGraph generator is a frequent offender when it comes to analysis jobs using multiple CPU.
- MadGraph is using multi-core mode by default:
  - To disable the multi-core mode the user needs to use `"/bin/generate_events --nb_core=1"`
- Gang qing (Glasgow) has done some tests to show C-groups limits this without affecting job length:



# Memory usage

- Increasing number of user analysis jobs requiring more resources.
- Shouldn't be unexpected as users are frequently running small scale production for validation.
- We are going to need to think about HIMEM (and MCORE) analysis queues.
- Must not do this in an ad hoc manner.
  - Setup Hospital queues in T1s?



# Pilot improvements

- Pilot not picking up cgroups correctly yet which would help (in future) with memory usage problems.
- Some analysis job errors because MAXWDIR and MAXINPUTSIZE are not set correctly in AGIS.
- Twiki states:
  - MAXINPUTSIZE - 14336 MB is the default (14 GB), but can be higher. Maximum copy-to-scratch input data total. Used in Panda Brokerage. To be obsolete once MAXWDIR is used in job brokerage in
  - MAXWDIR - Maximum available workdir space per job (MB). The initial value has been set as MAXINPUTSIZE + 2GB. Not used in Panda Brokerage yet. (APF-used)
- JEDI now uses schedconfig.maxwdir-2GB as input size.



# Suspicious files

- JIRA ticket tracking file recovery and suspicious file reporting:
  - <https://its.cern.ch/jira/browse/RUCIO-953>
- Can now declare lost files, but still not find suspicious files yet.
- Some (smaller) sites that have experienced data loss continue to cause problems for users.
  - Can we improve the clean up procedure?



# Rucio Download

- User trying to use Rucio download as part of his prun job.
- This should NOT be allowed.
- If users need access to data they should use panda which will ensure there job is sent to the data.
- Following up with user to see if there is a missing use case in prun.

