

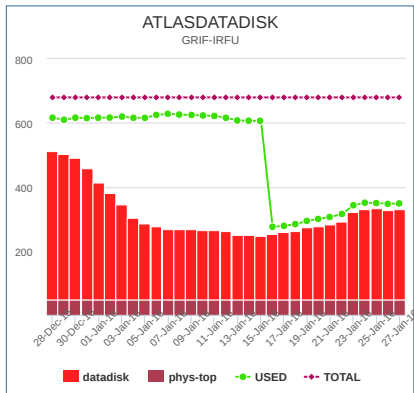
Consistency checks

Results and ongoing steps towards full automation

2016

- In normal conditions a file is both known to Rucio and present on the corresponding storage
- But errors happend:
 - Files deleted only from a storage but no from the DB: Lost.
 - Files copied to the storage but not registered on the DB.
- Ongoing effort to have automatic consistency checks on most sites.
- The tool to do it (rucio-auditor) is in place since a few months ago.
- This tool just give a list of suspicious files.

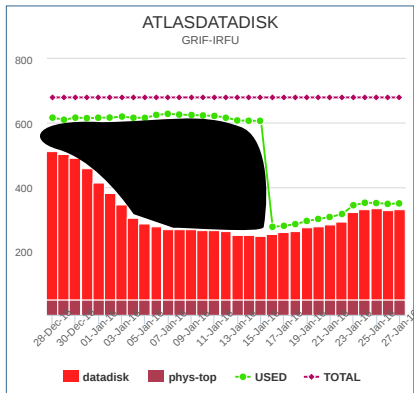
- Last incident between Christmas and the first days of January.
- Various sites full of dark data.
- Having dumps already available made possible to clean some endpoints without intervention of the admins.



Graphs on dashboard

<http://cern.ch/go/6SqM>

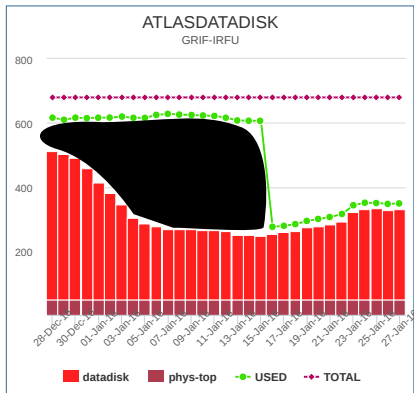
- Last incident between Christmas and the first days of January.
- Various sites full of dark data.
- Having dumps already available made possible to clean some endpoints without intervention of the admins.



Graphs on dashboard

<http://cern.ch/go/6SqM>

- Last incident between Christmas and the first days of January.
- Various sites full of dark data.
- Having dumps already available made possible to clean some endpoints without intervention of the admins.



Graphs on dashboard

<http://cern.ch/go/6SqM>

- Around 86 tickets solved, 1 unsolved. <http://cern.ch/go/kJL7>
- Around 51 sites are still implementing the generation of dumps.
- Cleaning still slow: Semi-automatic, various commands required (many possibilities for a typo or a human error).
- Dark data cleaning procedure:
 - Files detected as dark by rucio-auditor are verified in the DB.
 - If the file is not registered on Rucio it is deleted from the storage.
- Declaring lost files as bad replicas:
 - Files detected as lost by rucio-auditor are verified on the storage using GFAL.
 - If the files are not found on the storage, are marked bad.
- First steps towards a safe full automation of the cleaning of dark data.

- Around 86 tickets solved, 1 unsolved. <http://cern.ch/go/kJL7>
- Around 51 sites are still implementing the generation of dumps.
- Cleaning still slow: Semi-automatic, various commands required (many possibilities for a typo or a human error).
- Dark data cleaning procedure:
 - Files detected as dark by rucio-auditor are verified in the DB.
 - If the file is not registered on Rucio it is deleted from the storage.
- Declaring lost files as bad replicas:
 - Files detected as lost by rucio-auditor are verified on the storage using GFAL.
 - If the files are not found on the storage, are marked bad.
- First steps towards a safe full automation of the cleaning of dark data.

- Around 86 tickets solved, 1 unsolved. <http://cern.ch/go/kJL7>
- Around 51 sites are still implementing the generation of dumps.
- Cleaning still slow: Semi-automatic, various commands required (many possibilities for a typo or a human error).
- Dark data cleaning procedure:
 - Files detected as dark by rucio-auditor are verified in the DB.
 - If the file is not registered on Rucio it is deleted from the storage.
- Declaring lost files as bad replicas:
 - Files detected as lost by rucio-auditor are verified on the storage using GFAL.
 - If the files are not found on the storage, are marked bad.
- First steps towards a safe full automation of the cleaning of dark data.

- Around 86 tickets solved, 1 unsolved. <http://cern.ch/go/kJL7>
- Around 51 sites are still implementing the generation of dumps.
- Cleaning still slow: Semi-automatic, various commands required (many possibilities for a typo or a human error).
- Dark data cleaning procedure:
 - Files detected as dark by rucio-auditor are verified in the DB.
 - If the file is not registered on Rucio it is deleted from the storage.
- Declaring lost files as bad replicas:
 - Files detected as lost by rucio-auditor are verified on the storage using GFAL.
 - If the files are not found on the storage, are marked bad.
- First steps towards a safe full automation of the cleaning of dark data.

- Around 86 tickets solved, 1 unsolved. <http://cern.ch/go/kJL7>
- Around 51 sites are still implementing the generation of dumps.
- Cleaning still slow: Semi-automatic, various commands required (many possibilities for a typo or a human error).
- Dark data cleaning procedure:
 - Files detected as dark by rucio-auditor are verified in the DB.
 - If the file is not registered on Rucio it is deleted from the storage.
- Declaring lost files as bad replicas:
 - Files detected as lost by rucio-auditor are verified on the storage using GFAL.
 - If the files are not found on the storage, are marked bad.
- First steps towards a safe full automation of the cleaning of dark data.

- Around 86 tickets solved, 1 unsolved. <http://cern.ch/go/kJL7>
- Around 51 sites are still implementing the generation of dumps.
- Cleaning still slow: Semi-automatic, various commands required (many possibilities for a typo or a human error).
- Dark data cleaning procedure:
 - Files detected as dark by rucio-auditor are verified in the DB.
 - If the file is not registered on Rucio it is deleted from the storage.
- Declaring lost files as bad replicas:
 - Files detected as lost by rucio-auditor are verified on the storage using GFAL.
 - If the files are not found on the storage, are marked bad.
- First steps towards a safe full automation of the cleaning of dark data.

- Around 86 tickets solved, 1 unsolved. <http://cern.ch/go/kJL7>
- Around 51 sites are still implementing the generation of dumps.
- Cleaning still slow: Semi-automatic, various commands required (many possibilities for a typo or a human error).
- Dark data cleaning procedure:
 - Files detected as dark by rucio-auditor are verified in the DB.
 - If the file is not registered on Rucio it is deleted from the storage.
- Declaring lost files as bad replicas:
 - Files detected as lost by rucio-auditor are verified on the storage using GFAL.
 - If the files are not found on the storage, are marked bad.
- First steps towards a safe full automation of the cleaning of dark data.

- Around 86 tickets solved, 1 unsolved. <http://cern.ch/go/kJL7>
- Around 51 sites are still implementing the generation of dumps.
- Cleaning still slow: Semi-automatic, various commands required (many possibilities for a typo or a human error).
- Dark data cleaning procedure:
 - Files detected as dark by rucio-auditor are verified in the DB.
 - If the file is not registered on Rucio it is deleted from the storage.
- Declaring lost files as bad replicas:
 - Files detected as lost by rucio-auditor are verified on the storage using GFAL.
 - If the files are not found on the storage, are marked bad.
- First steps towards a safe full automation of the cleaning of dark data.

- Around 86 tickets solved, 1 unsolved. <http://cern.ch/go/kJL7>
- Around 51 sites are still implementing the generation of dumps.
- Cleaning still slow: Semi-automatic, various commands required (many possibilities for a typo or a human error).
- Dark data cleaning procedure:
 - Files detected as dark by rucio-auditor are verified in the DB.
 - If the file is not registered on Rucio it is deleted from the storage.
- Declaring lost files as bad replicas:
 - Files detected as lost by rucio-auditor are verified on the storage using GFAL.
 - If the files are not found on the storage, are marked bad.
- First steps towards a safe full automation of the cleaning of dark data.

- Around 86 tickets solved, 1 unsolved. <http://cern.ch/go/kJL7>
- Around 51 sites are still implementing the generation of dumps.
- Cleaning still slow: Semi-automatic, various commands required (many possibilities for a typo or a human error).
- Dark data cleaning procedure:
 - Files detected as dark by rucio-auditor are verified in the DB.
 - If the file is not registered on Rucio it is deleted from the storage.
- Declaring lost files as bad replicas:
 - Files detected as lost by rucio-auditor are verified on the storage using GFAL.
 - If the files are not found on the storage, are marked bad.
- First steps towards a safe full automation of the cleaning of dark data.

Common questions

- Who will handle the dump files (these are dark after all)
 - Dump files will be cleaned as part of the dark cleaning procedure.
- Dumps are slow to generate, how often should be generated?
 - We propose monthly (the auditor will test automatically once per month).
 - We can adapt to other frequencies (just let us know).
- In which date should be generated?
 - It can be any date, the auditor will take the newest based on the name of the dump file.

Common questions

- Who will handle the dump files (these are dark after all)
 - Dump files will be cleaned as part of the dark cleaning procedure.
- Dumps are slow to generate, how often should be generated?
 - We propose monthly (the auditor will test automatically once per month).
 - We can adapt to other frequencies (just let us know).
- In which date should be generated?
 - It can be any date, the auditor will take the newest based on the name of the dump file.

Common questions

- Who will handle the dump files (these are dark after all)
 - Dump files will be cleaned as part of the dark cleaning procedure.
- Dumps are slow to generate, how often should be generated?
 - We propose monthly (the auditor will test automatically once per month).
 - We can adapt to other frequencies (just let us know).
- In which date should be generated?
 - It can be any date, the auditor will take the newest based on the name of the dump file.

Common questions

- Who will handle the dump files (these are dark after all)
 - Dump files will be cleaned as part of the dark cleaning procedure.
- Dumps are slow to generate, how often should be generated?
 - We propose monthly (the auditor will test automatically once per month).
 - We can adapt to other frequencies (just let us know).
- In which date should be generated?
 - It can be any date, the auditor will take the newest based on the name of the dump file.

Common questions

- Who will handle the dump files (these are dark after all)
 - Dump files will be cleaned as part of the dark cleaning procedure.
- Dumps are slow to generate, how often should be generated?
 - We propose monthly (the auditor will test automatically once per month).
 - We can adapt to other frequencies (just let us know).
- In which date should be generated?
 - It can be any date, the auditor will take the newest based on the name of the dump file.

Common questions

- Who will handle the dump files (these are dark after all)
 - Dump files will be cleaned as part of the dark cleaning procedure.
- Dumps are slow to generate, how often should be generated?
 - We propose monthly (the auditor will test automatically once per month).
 - We can adapt to other frequencies (just let us know).
- In which date should be generated?
 - It can be any date, the auditor will take the newest based on the name of the dump file.

Common questions

- Who will handle the dump files (these are dark after all)
 - Dump files will be cleaned as part of the dark cleaning procedure.
- Dumps are slow to generate, how often should be generated?
 - We propose monthly (the auditor will test automatically once per month).
 - We can adapt to other frequencies (just let us know).
- In which date should be generated?
 - It can be any date, the auditor will take the newest based on the name of the dump file.

- Are the results available on some website.
 - Not yet.
 - After automation files unknown to Rucio will be shown on the usual place: https://rucio-ui.cern.ch/bad_replicas/summary.
 - There is a plan to have a table on the same site to display files detected as dark.
- What if there is any error in the dumps or the algorithm used is not 100% accurate?
 - Before taking action the state of the files will be checked with the Rucio DB as it is done currently with the manual cleaning.
 - I think it is less dangerous that having a person (me) doing the checks in a manual/semi-manual way.

Dark and lost data TWiki page <http://cern.ch/go/C9xr>

- Are the results available on some website.
 - Not yet.
 - After automation files unknown to Rucio will be shown on the usual place: https://rucio-ui.cern.ch/bad_replicas/summary.
 - There is a plan to have a table on the same site to display files detected as dark.
- What if there is any error in the dumps or the algorithm used is not 100% accurate?
 - Before taking action the state of the files will be checked with the Rucio DB as it is done currently with the manual cleaning.
 - I think it is less dangerous that having a person (me) doing the checks in a manual/semi-manual way.

Dark and lost data TWiki page <http://cern.ch/go/C9xr>

- Are the results available on some website.
 - Not yet.
 - After automation files unknown to Rucio will be shown on the usual place: https://rucio-ui.cern.ch/bad_replicas/summary.
 - There is a plan to have a table on the same site to display files detected as dark.
- What if there is any error in the dumps or the algorithm used is not 100% accurate?
 - Before taking action the state of the files will be checked with the Rucio DB as it is done currently with the manual cleaning.
 - I think it is less dangerous that having a person (me) doing the checks in a manual/semi-manual way.

Dark and lost data TWiki page <http://cern.ch/go/C9xr>

- Are the results available on some website.
 - Not yet.
 - After automation files unknown to Rucio will be shown on the usual place: https://rucio-ui.cern.ch/bad_replicas/summary.
 - There is a plan to have a table on the same site to display files detected as dark.
- What if there is any error in the dumps or the algorithm used is not 100% accurate?
 - Before taking action the state of the files will be checked with the Rucio DB as it is done currently with the manual cleaning.
 - I think it is less dangerous that having a person (me) doing the checks in a manual/semi-manual way.

Dark and lost data TWiki page <http://cern.ch/go/C9xr>

- Are the results available on some website.
 - Not yet.
 - After automation files unknown to Rucio will be shown on the usual place: https://rucio-ui.cern.ch/bad_replicas/summary.
 - There is a plan to have a table on the same site to display files detected as dark.
- What if there is any error in the dumps or the algorithm used is not 100% accurate?
 - Before taking action the state of the files will be checked with the Rucio DB as it is done currently with the manual cleaning.
 - I think it is less dangerous that having a person (me) doing the checks in a manual/semi-manual way.

Dark and lost data TWiki page <http://cern.ch/go/C9xr>

- Are the results available on some website.
 - Not yet.
 - After automation files unknown to Rucio will be shown on the usual place: https://rucio-ui.cern.ch/bad_replicas/summary.
 - There is a plan to have a table on the same site to display files detected as dark.
- What if there is any error in the dumps or the algorithm used is not 100% accurate?
 - Before taking action the state of the files will be checked with the Rucio DB as it is done currently with the manual cleaning.
 - I think it is less dangerous that having a person (me) doing the checks in a manual/semi-manual way.

Dark and lost data TWiki page <http://cern.ch/go/C9xr>

- Are the results available on some website.
 - Not yet.
 - After automation files unknown to Rucio will be shown on the usual place: https://rucio-ui.cern.ch/bad_replicas/summary.
 - There is a plan to have a table on the same site to display files detected as dark.
- What if there is any error in the dumps or the algorithm used is not 100% accurate?
 - Before taking action the state of the files will be checked with the Rucio DB as it is done currently with the manual cleaning.
 - I think it is less dangerous that having a person (me) doing the checks in a manual/semi-manual way.

Dark and lost data TWiki page <http://cern.ch/go/C9xr>