

# Why to Make Two Copies of Your Critical Archive Files

**NASA Center for Climate Simulation (NCCS)**

May 22, 2012

Ellen Salmon  
Code 606.2, GSFC  
[ellen.salmon@nasa.gov](mailto:ellen.salmon@nasa.gov)  
301-286-7705



## Single Archive Tape Copy Default Starting May 31, 2012



- Default becomes single tape copy of archive files:
  - ...for all newly created files *and*
  - ...all *existing* archive files.
- You can request a second tape copy of your **critical** archive files:
  - **dmtag -t 2 <archive\_file\_name>**



## Impact of Single Tape Copy Rather than Two Tape Copies for NCCS Archive



- A number of your archive files will take longer to retrieve from tape (temporary errors).
- “*A few more*” of your archive files will be unrecoverable (permanent errors) than with two copies.
- How many are “*a few more*”?
- Tape and hardware vendors haven’t been able to give us definitive answers, but NCCS’s recent “Get\_File” logs and tape media experiences can give us some ideas.



# Some Causes of Temporary and Permanent Tape Errors



- NCCS archive, May 1, 2012: ~150 million files, average file size ~100 MB, ~16 Petabytes total (single tape copy)
- **Temporary** errors: archive files available again within a few **hours** to a few **days**
  - Most common: tape drive problems, and robot problems in tape libraries
- **Permanent** errors: archive files **unrecoverable**, or possibly available after months (from recovery service)
  - **Archive tape problems** observed over the last two years:
    - Newest “T10KC” tapes, ~5 TB/cartridge, 2 detected bad of 3000 written (would average 100,000 unrecoverable files)
    - Previous generation “T10KA” tapes, ~2-500 GB/cartridge, 37 detected bad of 25,000 in use (would average 72,000 to 180,000 unrecoverable files)
    - *So over last 2 years, on average about one in 1,000 files could have been affected\* due to permanent media errors* (\*not all files may be affected, and recovery services may have been able to retrieve files)
  - Physical room/building problems that might affect a tape copy – rare, but possible
    - Fire (e.g., in on B28 roof April 2010, no fire detected)
    - Water or excess humidity – NCCS experiences over 20 years:
      - Excess humidity (alarms currently occur several times a year, *before* it harms media or hardware)
      - Severe roof leaks (observed so far: once in 20 years, ~30 tapes affected, data recovered after several months; tape libraries are no longer on top floors of buildings)



# NCCS Observed “Get\_File” Tape Operation Errors, Last 2 Years (May 2010 – May 2012)

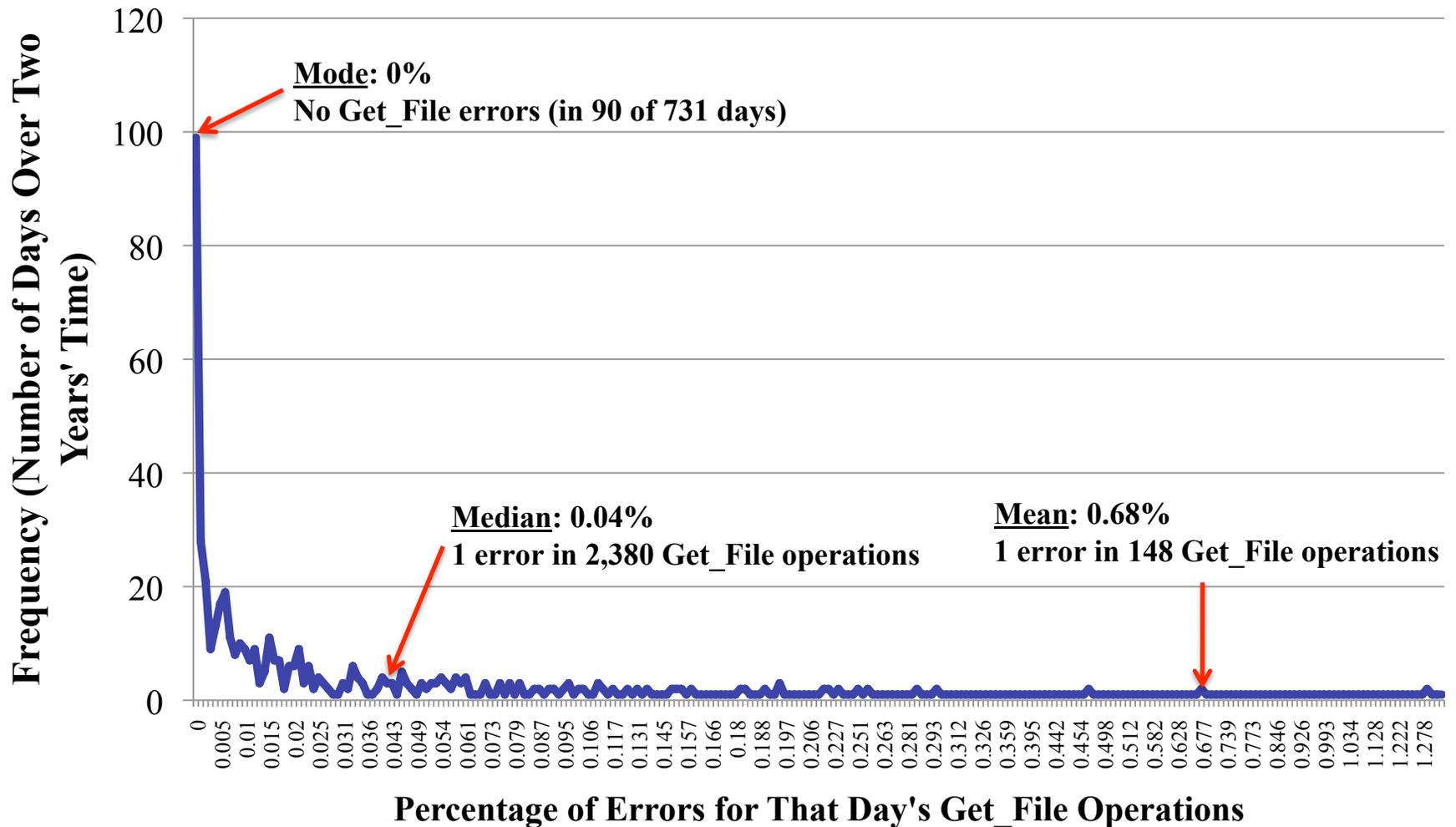


## Archive Tape Get\_File Errors from All Causes, Temporary and Permanent

What	Mode (most frequent value)	Mode Count (of 731 days total)	Mode % (of 731 days total)	Median	Mean	Standard Deviation	Two Standard Deviations Above Mean	Max.	Min.
Get_File Failure %	0%	90	12.3%	0.04%	0.68%	2.42%	5.52%	36.5%	0%
<i>One Failure in How Many Get_File Ops?</i>	<i>Zero</i>	<i>90</i>	<i>12.3%</i>	<i>2,380</i>	<i>148</i>		<i>18</i>	<i>3</i>	<i>Zero</i>



# NCCS Observed Percentage of “Get\_File” Tape Operation Errors, Temporary and Permanent, May 2010 – May 2012





## A Work in Progress—Caveats



- Things to take into account re error records reported to date:
  - When second tape copy has been available, NCCS has not been proactive about fixing first copy tapes promptly.
  - Vendors have not adequately documented error logs to help NCCS determine why a tape retrieve was obtained from the second copy.
  - *NCCS is improving its error recovery and tracking processes to accommodate the archive single tape copy environment.*



# Supporting Slide(s)





## NCCS Observed “Get\_File” Tape Operation Errors, Last 2 Years (May 2010 – May 2012)



- Days with no Get\_File errors: 90 (12.3%)
- Get\_File operations:
  - Mean, median, mode (count/percent), standard deviation, max, min
  - 130,179 : 73,000 : N.A. : 165,138 : 881,099 : 963
- Number of Get\_File errors
  - Mean, median, mode (count/percent), standard deviation, max
  - 511: 24 : 0 (90 count, 12.3%) : 2248 : 36,960
- Percentage of Get\_File operations with errors:
  - Mean, median, mode (count/percent), standard deviation, max
  - 0.676% : 0.042% : 0% (90 count, 12.3%) : 2.422% : 36.5%
  - 147.5 : 2380 : 0 : 41 : 2