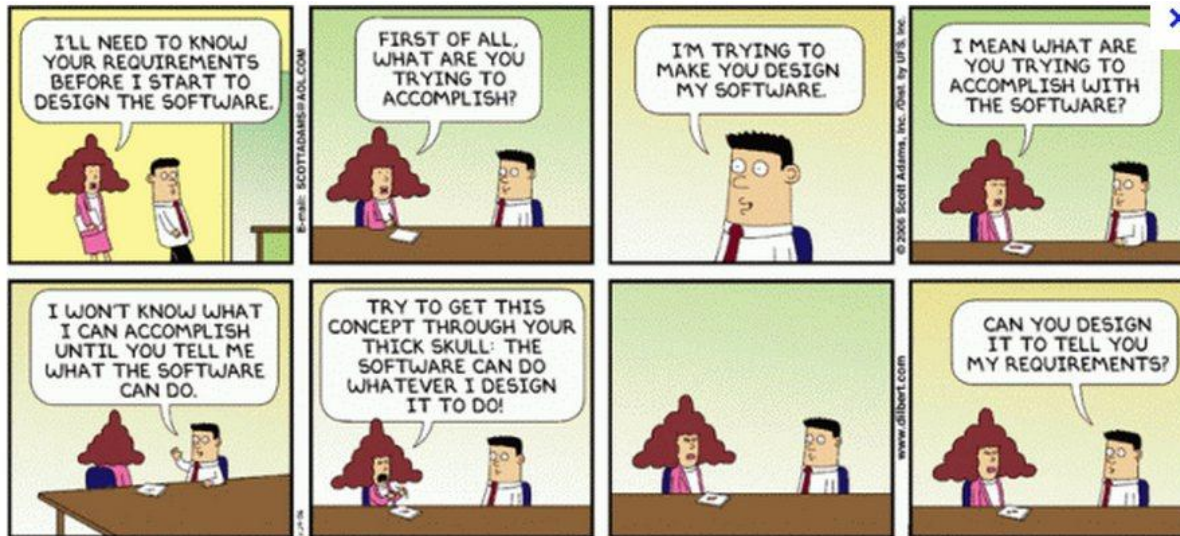


# ATLAS federated xrootd monitoring requirements

Rob Gardner

July 26, 2012



- We will need to iterate as we better understand how the federation is used
- Extend previous discussions from July 2011
- Attempt here to capture requirements from previous discussions
  - And input from Ilija Vukotic and Torre Wenaus (thanks!)
  - Will need to formalize within ATLAS so this is highly preliminary

# Recall July 2011 discussion

- Site level metrics identified
  - WAN direct read access related:
    - MB/s read
    - # remote connections
  - File caching related (FRM e.g.)
    - MB/s into a site
    - # success and # failed transfers
    - # active movers
- Aggregate locally and publishing to central collector for federation-level display

We will need to extend this for the production infrastructure

# General federation monitoring wishes

- Site-level metrics as well as aggregate, federation-level metrics useful for assessing both functional status and performance
- Redirection statistics: fraction of time accesses are local, redirected within a region, cloud or global
- From a job and data management systems perspective we'll need deeper information to generate a profile for federated access patterns to a specific site, and collections of sites

# Capturing a list

- Site availability and redirection functionality
  - SSB-like (Jarka) and SLS (central services)
- Aggregate federation IO accesses and by site
  - With time histories
  - Redirection rates
  - Authentication successes/failures
- Global, cloud and regional aggregate summaries for federated IO
- Number of files opened
  - Distinguish direct access versus copy
  - Distinguish local versus WAN
- IO rates
  - Distinguish direct access versus copy (hard!)
  - Distinguish local versus WAN (also hard!)

# List, cont

- Statistics for files actually used and mode of access
- User statistics for direct access versus copy
- Viewable as current, real-time snapshots and as archival for time histories
- For brokerage, link with Ilija's cost matrix – “click down” to get the story behind the cost

# List, cont

- plots ranking sites by data (file counts, byte counts, user counts) served/consumed (copy and direct)
- plots ranking sites by availability, reliability
- file lifetime distributions by site
- "active" data volume at a site, absolute and as fraction of capacity, where "active" file is one used in the last X weeks/months
- fraction of file opens that find a copy local to the site vs. having to open/retrieve a remote copy (redirection statistics)
- plot of file age at deletion (cleanup), and plot of avg file age at deletion by site

# Summary

- In the next weeks we will better formalize something official to work from after consulting more folks within ATLAS ADC
- In the meantime we can focus on the obvious basic metrics