Federating Storage NDGF Style

When does it make sense? And when doesn't it?



NordForsk

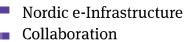
Nordic e-Infrastructure Collaboration



Overview

- The need for storage consolidation
- How NDGF is consolidated
- Boundary conditions necessary
- When would you look for different solutions
- How WLCG can help to facilitate this

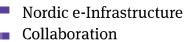




The need for storage consolidation

- Well covered in the WLCG workshop
- Storage is stateful
- Requires non-trivial manpower to run well
 - And keep up with changing user requirements
 - Dark data, low performance clogging transfer slots, downtimes
- Benefits from economy of scale
 - Especially with regards to manpower
- Experiment effort is disproportionally spent on small SEs
 - Signalling that "small" SEs aren't worth the effort





How NDGF came to be federated

- Once upon a time...
 - Wanted a Nordic Tier-1
 - Meant a single storage endpoint
 - No single country was big enough to fund one
 - Money could not be reasonably sent across borders
- Solution
 - Single namespace and dCache head nodes
 - Storage pools distributed between sites with local admins just handling servers, tape libraries, raidsets, and occasionally upgrading/restarting dCache according to directions



Federated dCache NDGF-Style

- Not only in the Nordics:
- ATLAS Great Lakes Tier-2 also runs in a similar fashion
 - Some minor technical details differ, but the idea is the same
- SI-SIGNET-T2 (at IJS in Slovenia) joined srm.ndgf.org
 - Main motivation was that tier-1 space is more valuable to ATLAS than tier-2, and since we're distributed, "why not?"
 - Probably makes local effort easier too, compared to running a fully independent storage element
 - Roughly same story for SE-SWEGRID-T2



Boundary conditions necessary

- You can't send all the money to one site and have a big SE
 - If you can, you're probably better off doing that from a cost efficiency point of view
- You're fine with one central site getting most of the visibility
 - A site that runs the storage as pools only for another site is going to look smaller than if it ran it's own SE
 - Even if that separate SE would be too small to be of much value to the users
- You have sufficient internal networking
 - For Close-SE direct file access to work, AGLT2 makes temporary copies of all files accessed on the close set of pools



Boundary conditions necessary

- A good team for the central SE
 - A little bit more work
 - A little bit more corner cases
 - Arguably you'd need that anyway for delivering good storage
- A will to deliver the most bang for the buck to users
 - Listen to the experiments on where they find value, i.e. large wellrun storage elements
 - And willing to change to maximize benefit



When would you look for different solutions

- If you can centralize the money
 - Distribution comes with some overhead in manpower etc
 - Even if there are some benefits to the distributed nature, on the whole, you probably are better off with the classic solution
- If you can't cooperate
 - Central site and pool sites need to get along
 - Define a clear boundary between central and site responsibilities
- If visibility is important
 - "But I've pledged 200TB of storage, it'll look like I'm not delivering according to pledges!"



How WLCG can help

- The pledge visibility issue
 - Does SI-SIGNET-T2 deliver storage according to pledges?
 - Well, yes, if you take the Slovenian-sized chunk out of the installed capacity of NDGF-T1's srm.ndgf.org it matches pledges
 - But then, does NDGF-T1 deliver storage according to pledges?
 - A good strategy for how to handle this would be nice for cross-site consolidation
 - Since one of the big use cases is for tier-1s to gobble up storage from tier-2s, this is a highly relevant issue
- Sharing knowledge of how to do this



Nordic e-InfrastructureCollaboration

Questions?

