T3g software services

Outline of the T3g Components

R. Yoshida (ANL)

Outline

- Introduction
- Basic services: NFS, user account management
- Software installation, maintenance and user interface: ATLASLocalRootBase and manageTier3SW packages.
- CVMFS (CERN Virtual Machine File System): providing Athena, database releases, and conditions data.
- Batch system: Condor
- Analysis data management: XrootD
- Local batch parallel processing interface: ArCond
- Panda for Tier3
- Getting ATLAS data from the Grid

Introduction

- What's presented at this meeting is a coherent set of software services for a T3g.
- What's needed:
  - Provide all needed capabilities for T3g.
  - Must be relatively easy to configure and install.
  - Must be relatively easy to maintain.
- Take advantage of ATLAS-wide (and beyond) resources:
  - Done in concert with all of ATLAS (not just US ATLAS)
  - Conform to the conclusions of 3-months ATLAS T3 working groups which are just concluding.
- Certainly not the only choices that could be made—however
  - Standardization will help you get help!
  - Standardization will lessen the support load.
  - These particular choices have been tested for some time. Building up another set will be a significant amount of work.

Basic Services

- User management: LDAP
  - We considered using simple password+shadow files.
  - Rejected because too cumbersome (and too easy to make a mistake) for even a small cluster.
  - LDAP is widely used, more modern than NIS. Easier to deal with groups.
- Integration with existing user-base.
  - University of department user database
  - This is being done here at ANL ASC.
  - If you can do standalone, this is a lot easier!
- Shared file system: NSF4
  - Shared software
  - User area (can mount a separate area from a different cluster—being done here)

ATLAS and Grid
Software installation and management

- Integrated package called ATLASLocalRootBase

- Installed and updated by manageTier3SW

- Takes care of a lot of non-system software

  - Clients to interact with the grid

  - Standalone Root

  - C++ compilers

  - etc.

- Controlled and automatic updates

- Unified and easy to use user interface

Athena, Database versions, conditions data

- CVMFS (Cern VM File System)
  - Part of Cern VM Project
  - CVMFS itself, however, doesn't need to run on VM
  - It's a web file system
  - Athena, database versions and conditions data are updated on the server
  - Frees T3g admins from having to update these.
  - Will need to work in conjunction with a local squid (web caching).
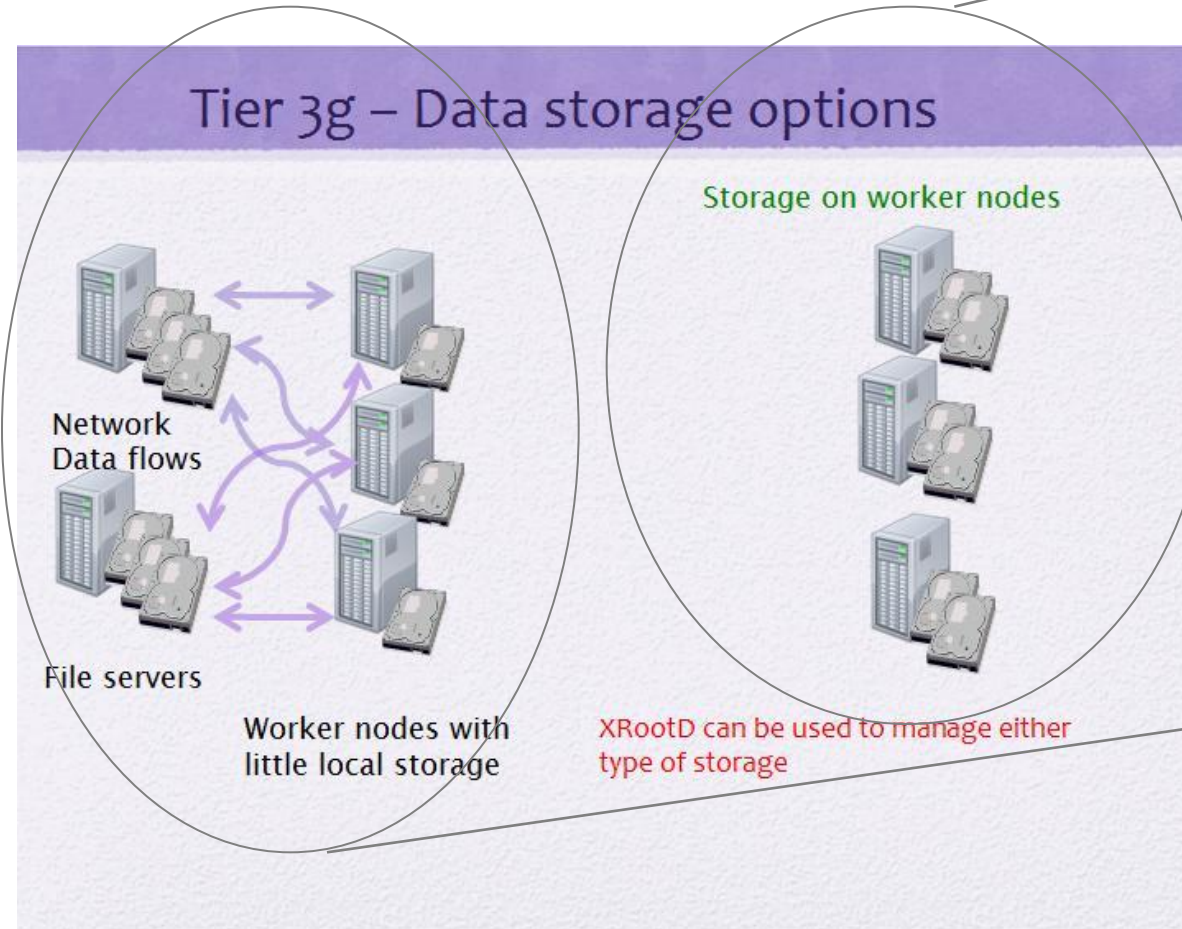
Batch system

- CONDOR

- Widely used stable system
- A strong support for the ATLAS T3 project from the CONDOR team
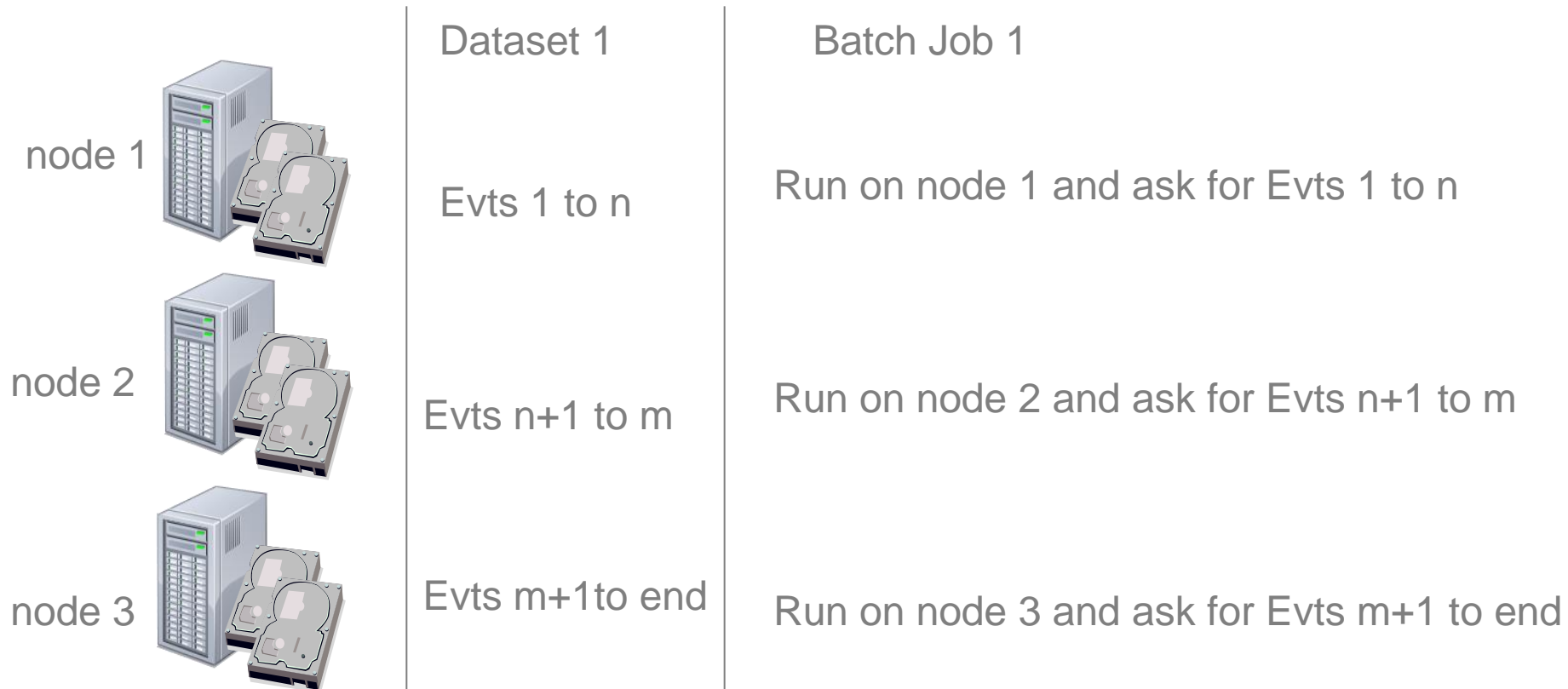- Currently not considering other batch systems.

Batch system

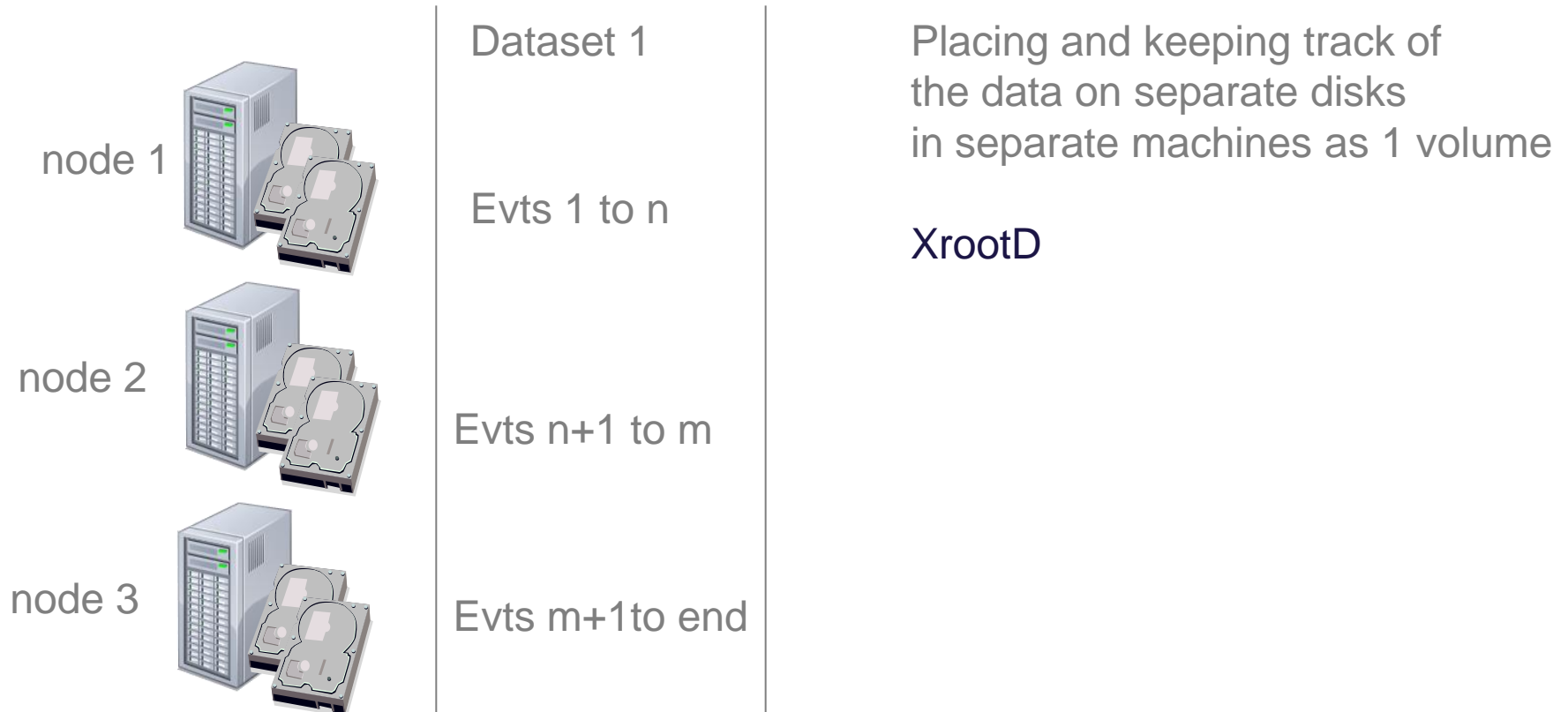Data located with the processor. Has good performance. This is the current baseline

Running directly off data on NFS has severe limitations. Needs a "data mover".

Handling the data in the distributed disks

	Dataset 1	Batch Job 1
node 1	Evts 1 to n	Run on node 1 and ask for Evts 1 to n
node 2	Evts n+1 to m	Run on node 2 and ask for Evts n+1 to m
node 3	Evts m+1to end	Run on node 3 and ask for Evts m+1 to end

Local CONDOR parallel processing interface: ArCond

8 June 2010

Handling the data in the distributed disks

node 1

node 2

node 3

Dataset 1

Evts 1 to n

Evts n+1 to m

Evts m+1to end

Placing and keeping track of
the data on separate disks
in separate machines as 1 volume

XrootD

T3 PANDA

- Use PANDA server at CERN to submit to local batch queue

- Same interface as usual Pathena

- Need to supply a list of files. (Panda can't know about the files in your local cluster)

- For now, the jobs are not submitted with the data location considered.

- The "data affinity" is planned as a next step.

Getting ATLAS data from the Grid

- We have three solutions, none of them very satisfactory
1. dq2-get
   - OK for small data sets.
   - User initiated—needs extensive baby sitting for larger data sets.
   - No control at the data source end—potential to overwhelm Storage Element at T2 or T1.
2. Grid Storage Element (SRM)
   - Data transfer is much more reliable.
   - Can have data subscription.
   - But.. you are now a part of the Grid Storage. This is a bit like being a disk in a RAID array.   You can't do what you want with the data, and if you're not behaving correctly, you can be kicked out.
3. dq2-get/FTS
   - This is like dq2-get but the request is queued at the end point T1 or T2.
   - No (or considerably less) baby sitting.
   - Controlled at the source so safer for T1 and T2.
   - Just becoming available.
   - But.. currently causes large problems at T1 or T2 if a T3 requesting files stops accepting data.

8 June 2010

Getting data from the Grid

- Currently we cannot recommend installation of an SRM or dq2-get/FTS.

- Stick with normal dq2-get if you want to put up a T3 in the near future.

- We'll be testing dq2-get/FTS in the near future.

- Strategy of how to deal with the problem—July time scale.

8 June 2010