PetaShare Tutorial

04/15/2009

Ismail Akturk & Sreekanth Pothanis
Outline

• Statement of problem
• PetaShare Overview
  – Architecture
  – Client tools
  – Demo
Research Problem

• Scientists/Researchers who deal with **large-scale** data need to share data with their colleagues for the sake of collaboration.

  – Challenges
    • Different environments
      – Operating Systems, Architectures
    • Security
      – Authentication and Authorization
    • Network
      – They are geographically apart from each other

• Current filesystems may scale and work in LAN, but not in WAN
An Approach

- An **infrastructure** on top of variety of underlying systems (filesystems, OSs, etc)
  - Provides transparent handling of low-level data management issues
  - Lightweight access tools
  - Based on scalable server/client model
Let's Name it
What is it and not?

• PetaShare is **NOT**
  – Filesystem
  – Stand-alone Software

• PetaShare is
  – Storage Management Infrastructure
    • That provides filesystem-like infrastructure
      – Accessing to PetaShare is convenient as much as accessing to local filesystem.
    – Composed of variety of tools...
PetaShare

• Provides
  – Global Namespace among distributed resources
  – Interfaces to access distributed resources from variety of OSs and architectures.
    – Pcommands & Windows iCommands
    – Petashell
    – Petafs
    – Windows Browser
    – Web Portal
  – Meta-data management interface*
    • Search functionality through Web Portal

(*)Out of scope of this tutorial
Facts

- Spans among six Louisiana research institutions
  - Seventh site (LaTech) is under way :)
- Manages
  - 300TB of disk storage
  - 400TB of tape storage (will be online soon)
- Leverages LONI
  - 40Gbps
Architecture
A simple Illustration
Research Group: **RG1**

- userA
- userB
- userC

Global Namespace:

/tempZone/home/RG1

- file1
- file2
- file3

Diagram:

- LSU
- UNO
- TULANE
- ULL
- LSUS
- LSUHSC

- File1
- File2
- File3

- Get file3
- Get file1
- Get file2
How to Access PetaShare Resources

• Through Client Tools
  – Pcommands
  – PetaShell
  – PetaFs
  – Web Portal
  – Windows Browser
Which Tool Should I Use?

- **Petafs**
  - Virtual File System
    - Allows you access PetaShare resources through your existing filesystem by mounting PetaShare resources
    - Requires FUSE support in kernel, installation of FUSE requires root privileges
    - Available in Linux
      - Mac version will come up shortly
• Petashell
  – It is similar to petafs in a way that it allows you access PetaShare resources as they are extension of your existing filesystem.
    • Only within petashell
  – Do not require any root privileges
  – It is an alternative to petafs in case you do not have FUSE installed in your system.
  – Available in Linux
• Pcommands
  – Most portable client tool
  – Special commands to access PetaShare
  – Provide advance utilities to manage low level data issues such as replication, physical movement of data, etc...
  – Available in Mac, Linux
• iCommands for Windows
  – Similar to Pcommands.
  – Use full for upload and download of large files in windows.
  – Can be used in Batch processing.
Windows Browser

- User friendly
- No installation required
- High upload/download speeds
- Can be downloaded from downloads page of www.petashare.org
• Petashare Portal
  - Accessible at http://dsl-yoda.csc.lsu.edu/petashare
  - No installation required
  - Compatible with all class A browsers
  - Supports search features
  - Provides html links to files on Petashare
Demo

- Pcommands & Windows iCommands
- Petashell
- Petafs
- Windows Browser
- Web Portal
How to transfer your bulk data into PetaShare?

Stork Tutorial

same place/time

on April 27th
Questions