dCache.org

Federated File System

(let's learn from the others' ~~mistakes~~ expirience)

Tigran Mkrtchyan for dCache Team

Chuck Lever chuck.lever@oracle.com
Robert Thurlow robert.thurlow@oracle.com

Fermilab NDGF Nordic DataGrid Facility LSDMA DESY HELMHOLTZ ASSOCIATION

Déjà vu

Internet Engineering Task Force (IETF)
Request for Comments: 5716
Category: Informational
ISSN: 2070-1721

J. Lentini
C. Everhart
NetApp
D. Ellard
BBN Technologies
R. Tewari
M. Naik
IBM Almaden

January 2010

Requirements for Federated File Systems

Abstract

This document describes and lists the functional requirements of a federated file system and defines related terms.

What this presentation about

- Show industry attempt to solve similar issue
- Show ideas and implementation details
- Show weak points and problems
- Help to detect and avoid “well know” mistakes

Requirements (join effort SUN + NetApp)

- Provide a set of protocols to turn a collection of fileservers into a federation
- Provided namespace hosted on different fileservers
- Fileservers can belong and managed by different administrative entities

What is FedFS

- A way to build a uniform namespace
  - Do you still remember AFS?
  - Make use of existing (unmodified) servers
  - Make use of existing (unmodified) clients
- Keep POSIX semantics
- Migrate files to new locations
- Replica list, geo/load balancing

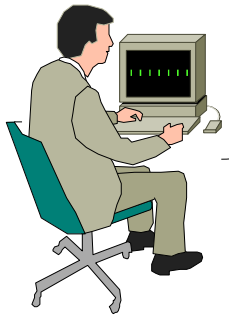
Junction/Referral

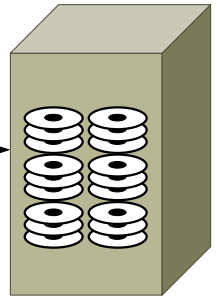
- **Junction:** A filesystem object used to link a directory name in the current filesystem with a directory on a different filesystem.
  - A special ***SYMLINK***-like object with only one attribute – filesystem location information
  - Querying for any other attr will return ERR_MOVED
- **Referral:** response with location information returned to the client.

Implementation Details

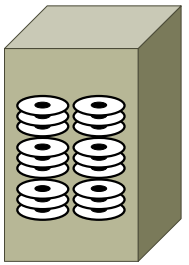
/wlcg/atlas/dataset1/file1

cd /wlcg; cd atlas

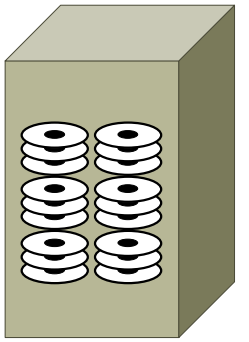




Server A



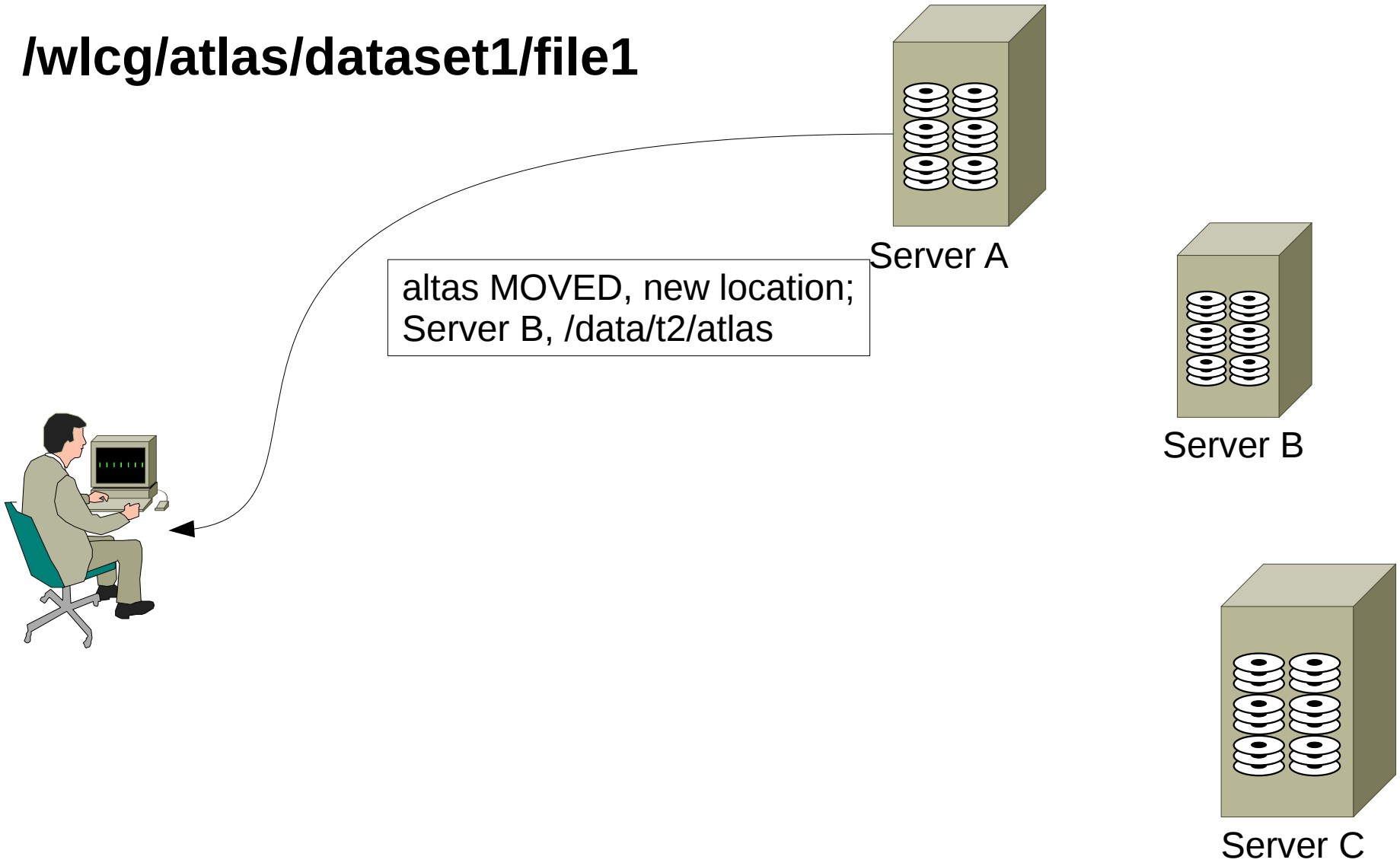
Server B



Server C

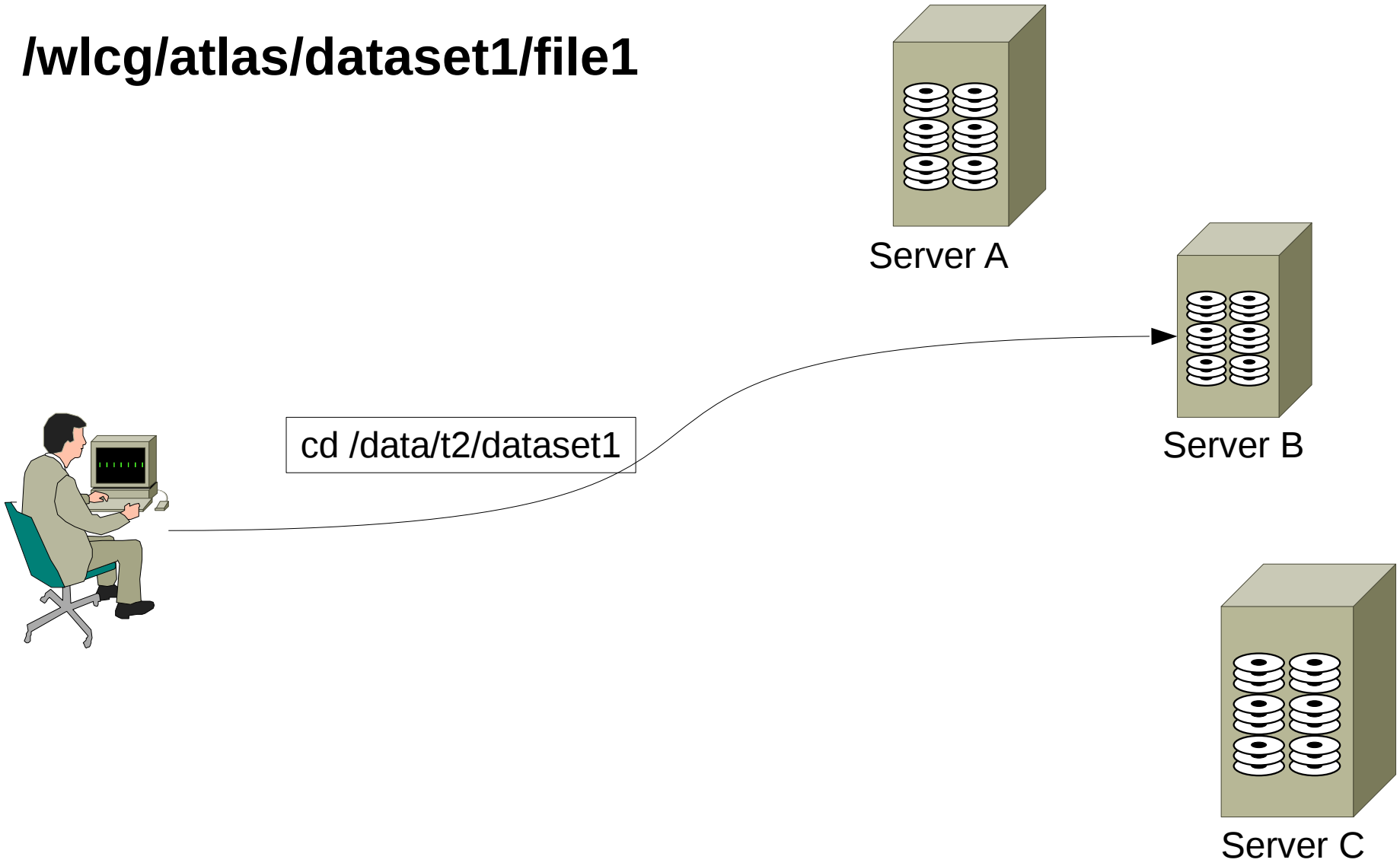
Implementation Details

/wlcg/atlas/dataset1/file1



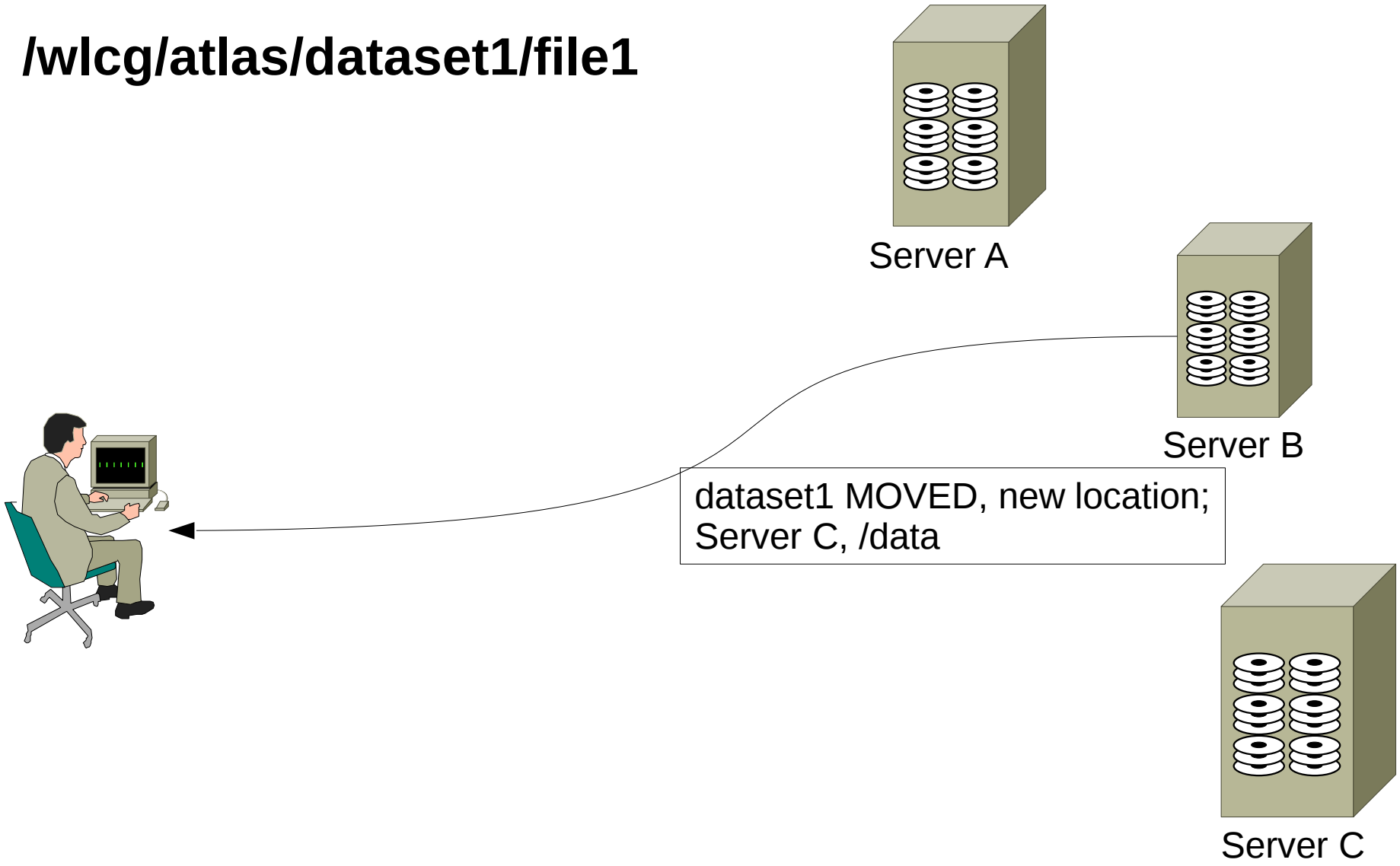
Implementation Details

/wlcg/atlas/dataset1/file1



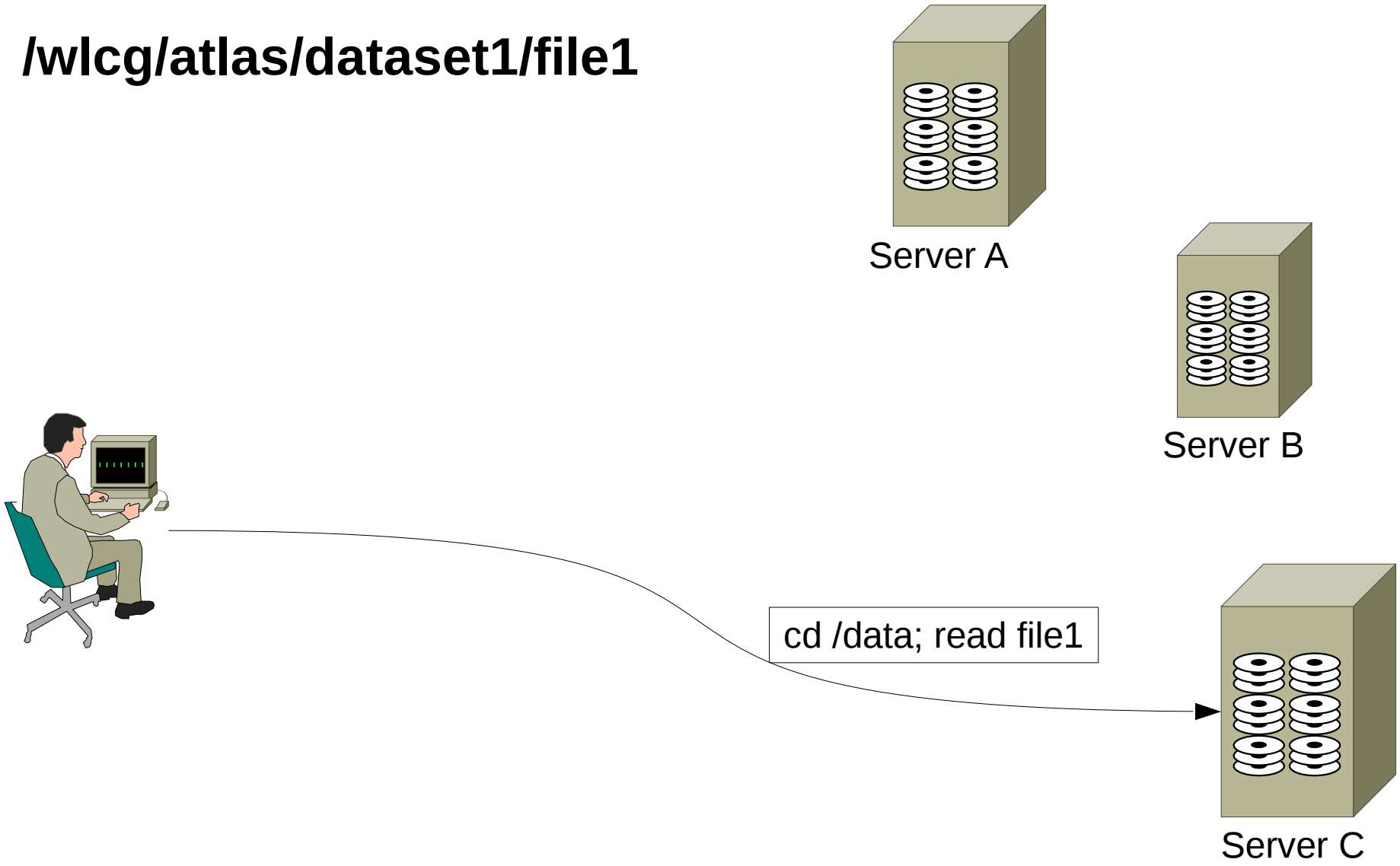
Implementation Details

/wlcg/atlas/dataset1/file1



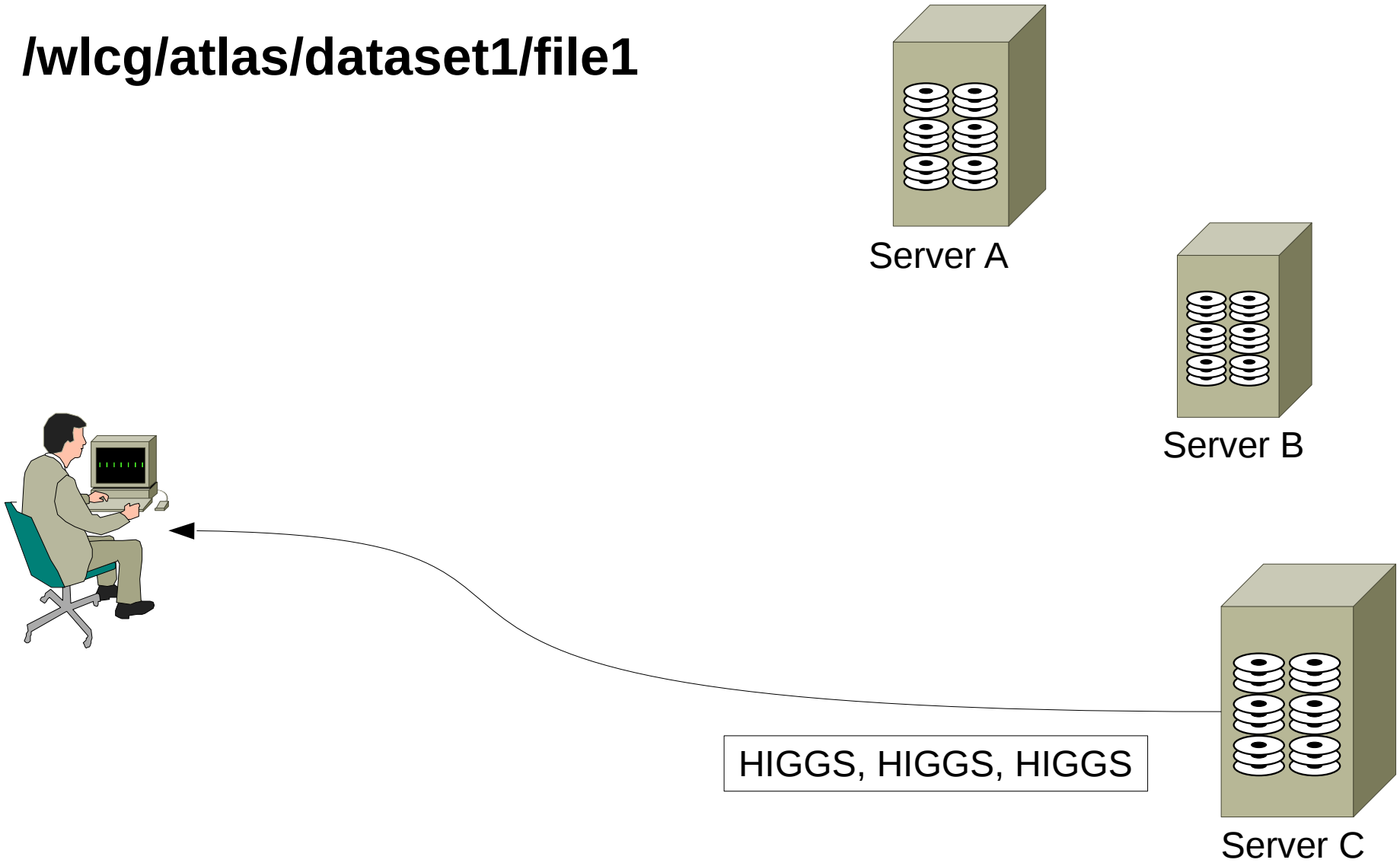
Implementation Details

/wlcg/atlas/dataset1/file1



Implementation Details

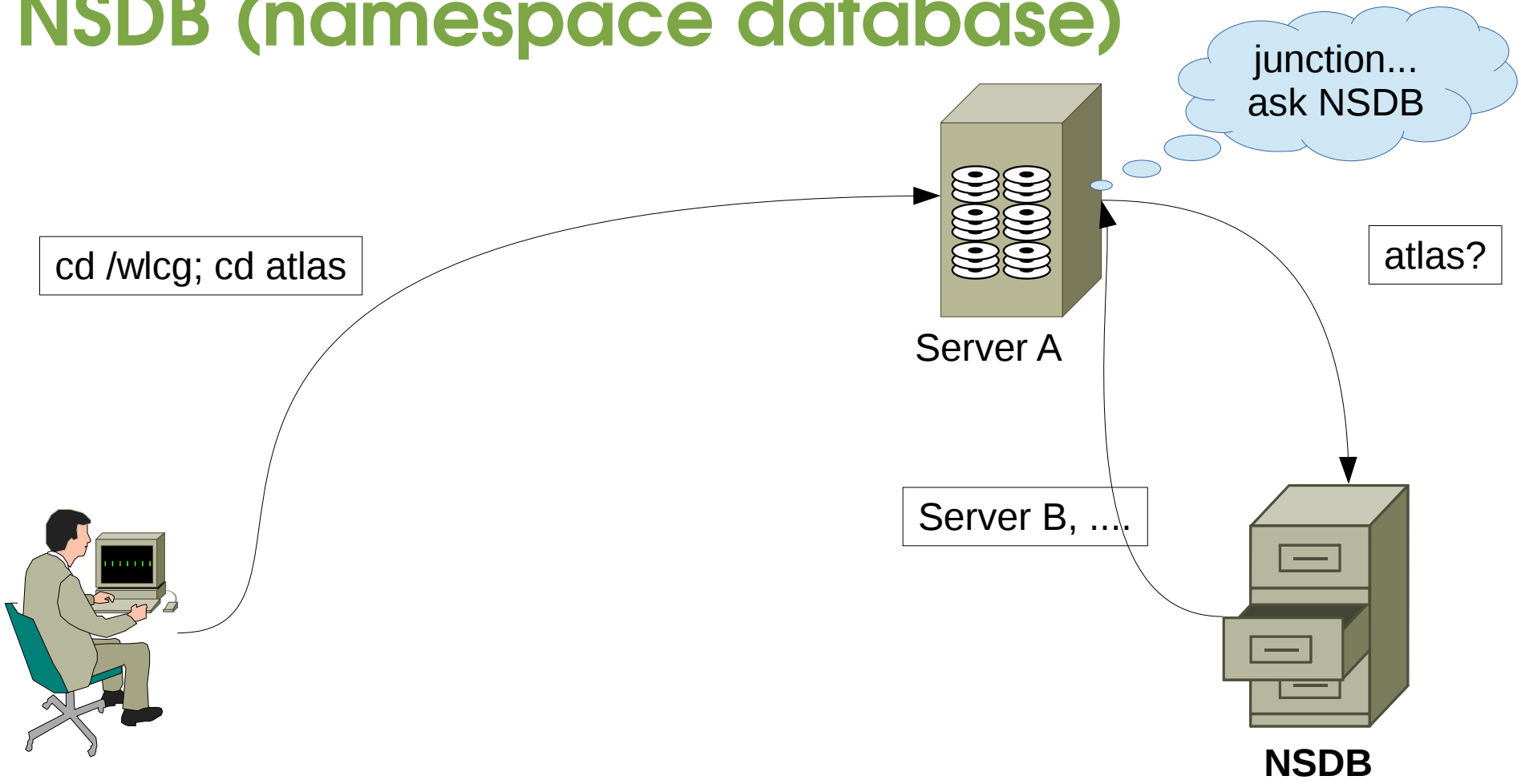
/wlcg/atlas/dataset1/file1



Transparent

- Transparent for end-user
- **Server-C** an aware of re-directions
  - Only Servers **A** ans **B** must provide referrals

NSDB (namespace database)



NSDB

- slowly changing information
  - easy to replicate
- small set of information
  - easy to manage / replicate
- can be managed locally
  - each server/site may have it own NSDB
- RFC7533: Admin Protocol for Federated File Systems
  - protocol to inject junctions points into FS
  - Solaris and Linux implementations
  - NFS only :(

Good sides (linux)

- Available in EPEL repo
- Implemented as automounter + LDAP
- Uses DNS service record
  - no mount required
  - /nfs4/domain.com (do you remember AFS?)
- NFSv4 compliant client is sufficient
  - linux, solaris, OSX 10.10 (?)
- 'last server' can be any valid V4 server

Down sides of FedFS

- Protocol (semantics) specific
  - All server MUST talk the same protocol
- LDAP specific
- Assumes user can access any server
  - Federated identity required
- Management protocol does not define access controls
- No production installations

More info

- https://tools.ietf.org/rfc/rfc5716.txt
  - Requirements for Federated File Systems
- https://tools.ietf.org/rfc/rfc7533.txt
  - Administration Protocol for Federated File Systems
- https://tools.ietf.org/id/draft-adamson-nfsv4-multi-domain-federated-fs-reqs-02.txt

Implementation Details