# Easing the Burdens of HPC File Management

**Stephanie Jones**, Christina Strong, Aleatha Parker-Wood,
Alexandra Holloway, Darrell D. E. Long

UC Santa Cruz

SSRC Storage Systems Research Center
U.S. Department of Energy
Office of Science
iTRIS Center for Information Technology Research in the Interest of Society
Baskin Engineering UC Santa Cruz

# Introduction

- Scientists waste time on manual metadata management
  - Manual correlation of related files and data
  - Remembering which storage system data is on
  - Lengthy file names with human error
  - Time spent searching for the right data

- We propose a provenance enabled file system to prevent this
  - Provenance integrated with traditional metadata
  - Unified search space over primary and archival storage
  - Generate descriptive names for files
  - Query results ranked by importance to the scientist

# Introduction

- Interviewed scientists from LANL, PNNL and NOTUR
- File system use is not always “correct”
  - Some scientists store everything on a parallel file system
  - Others only keep static files on NFS
- Results in
  - Degradation of file system performance
    - Managing loads not designed to handle
  - Scientists managing their own metadata information

# File Organization Methods

- Physical copies
  - Handwritten metadata information in notebooks
  - Print outs of typed metadata information in binders
- Electronic copies
  - PowerPoint presentations
  - Spreadsheets
  - Encode metadata into file or directory names
  - Text files
- Scientists feel the file system is lacking
  - Not recording the information needed
  - Easier to manage own data than find it again

# An Ideal System

- Scientists don't want to worry about where their data is stored
  - Care that it *is* stored, and that they can get it back
- Query metadata specific to their experiments
  - Coordinates with a specific temperature
- Create tags to classify and search their data
  - Tag all files used in an experiment
- Relationships between files are created automatically
  - Visualization file came from this output data

# Background

- Data provenance
  - Content based
  - Information flow
- Workflow provenance
  - Prospective
  - Retrospective
- File system ranking
  - Use existing metadata
  - Create new metadata by inferring links between files
  - Manually specify relationships between files

# Provenance Enabled File System

- Data provenance will track file relationships for scientists
- Provenance can be used to create a unified search space over multiple storage systems
- Provenance and traditional metadata can be used to create descriptive names for files
- Rank search results to more accurately return useful files from queries

# Provenance in HPC

- No known implementations of a provenance enabled file system in HPC
- However, provenance has a strong presence in grid computing
- Our work assumes provenance is collected and available

# Storing Provenance

- Common method for storing provenance is in a database
  - MySQL
  - Berkeley DB
- Traditional DBMS are often a poor solution for search and indexing applications
  - Customized, application-specific performs better
  - Databases often optimized for read or write
- We are exploring ways to add provenance to Ceph
  - Store provenance in the metadata server
  - Embed provenance data in files

# Unified Search Space

- LANL has 3 different storage systems
  - PFS, NFS, HPSS
- Querying each can be time consuming
- Create a unified search space
  - A query must be able to access all types of metadata
  - A query must be able to span all types of storage
- Use transient provenance
  - Creates a record of archived data on primary storage

# Transient Provenance

- Extension of information flow provenance
- Tracks flow of data off a provenance aware system
  - Designed to help identify potential data leaks
  - Keeps a metadata record of data that leaves
- Archiving moves data off of primary storage
  - Query over primary storage will include the metadata records of archived data
- Currently covers NFS and HPSS

# File Naming

- Scientists often encode experiment parameters in file or directory names
  - libego_alpha1_175_old_formula_uct2_final

- Can result in many different problems
  - What was alpha?
  - Did I do old_formula or formula_old?
  - I can use 1024 characters on this system, but only 256 on that one

# Generating File Names

- Current issue
  - Multiple files with the same name
  - Low entropy
- Obvious Solution
  - Use unique identifier for file names (i-node number)
  - High entropy
- Better solution
  - Determine attributes that are both unique and interesting

# Choosing Attributes

- Currently experimenting with techniques for creating meaningful file names
- Techniques from
  - linguistics (Zipf's law)
  - faceted search
- Data created by human behavior often follows Zipfian distributions
- Data created by automatic processes is often random or uniform
- Distribution of data will aid in identifying meaningful attributes for file naming

# Preliminary Results

- Naming MP3s
  - Zipf's Law, high entropy, random
  - top three attributes
- Zipf's Law
  - 'Robert Kraft-Swing Kids-2008-06-18T06:23:53Z'
  - Artist - Album - Date Added
- High Entropy
  - 'AF080E109902BB43-6895-file://localhost/Users/aleatha/Music/iTunes/iTunes%20Music/Robert%20Kraft/Swing%20Kids/01%20Sing,%20Sing,%20Sing%20(with%20a%20Swing).mp3'
  - Persistent ID - Track ID - Location
- Random
  - '6895-Robert Kraft-MPEG audio file'
  - Track ID - Artist - File type

# Query Result Ranking

- With the massive amounts of data being generated, need a way to find important data quickly
- Naming files with uniquely identifying and interesting information is half
- Presenting data that is most important (to the scientist) first
  - Think about a search on Google

# Why can't we just use Google?

- Google relies on the innate structure of the web
  - Links between pages are implicit endorsements
- Current file system structures:
  - Directories?
  - Hard/soft links?
  - File names?
  - Access times?
- None of these are actually endorsements

- Without this, Google and other modern search engines are <u>just another similarity search</u>

# Provenance and Search

- Provenance tracks data flow, file opens, file closes
- In combination with other metadata, can tell us:
  - How many people used a file
  - How recently
  - How often
- Provenance can be used as an endorsement
  - What files people use
  - How often they use those files

# Leveraging PageRank

- By interpreting provenance as endorsement, we can **leverage** existing ranking algorithms like PageRank, and **create** new ones
- How do we use the provenance graph to determine usefulness?
- PageRank uses links to determine the probability of spending time on a page
- Can do the same, but with some modifications

# PageRank

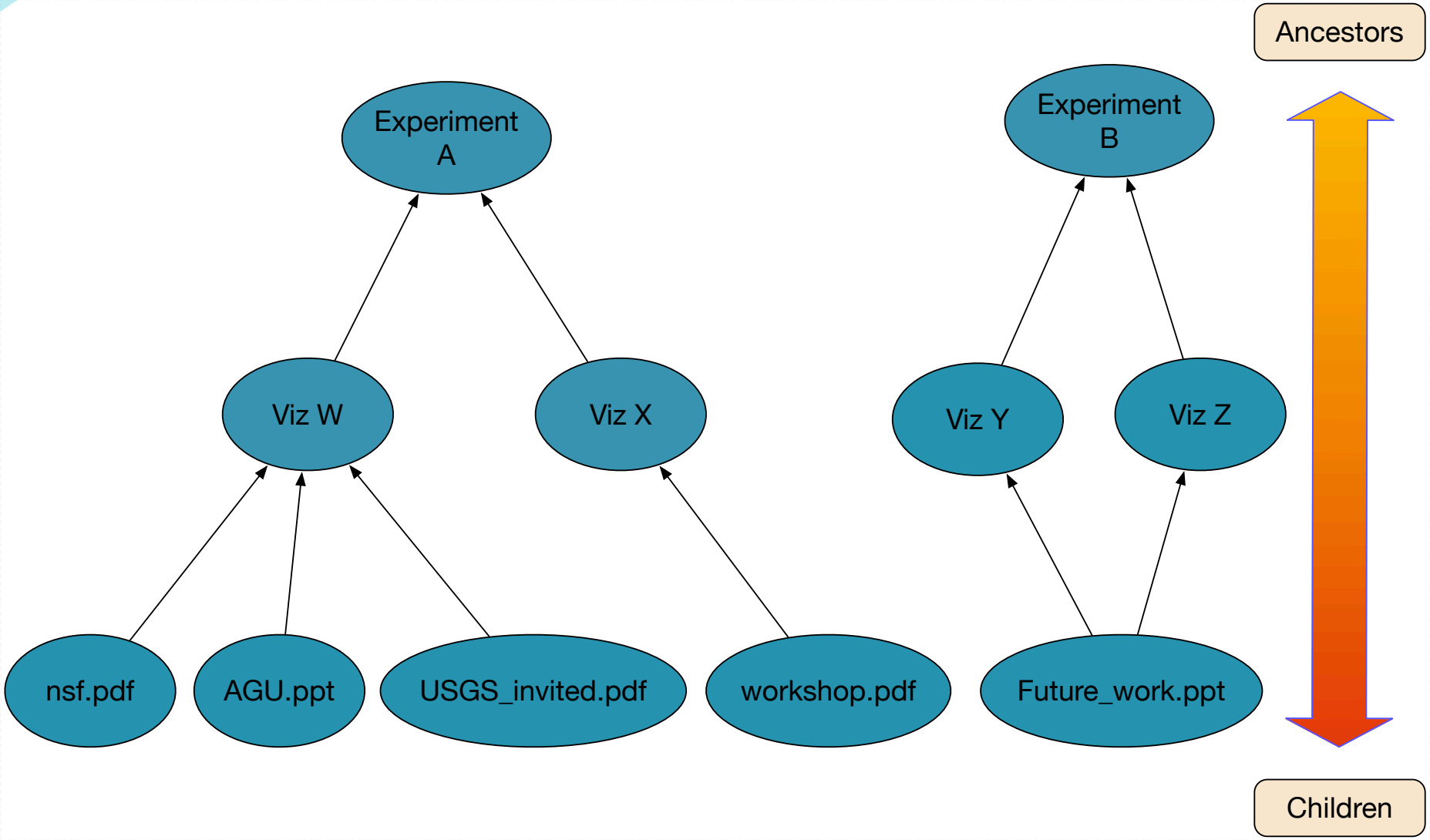
- Stationary distribution of the Markov chain described by the transition matrix of the web graph
- What is the probability of arriving at a given page?

- An infinite number of web surfers randomly click links for an infinite amount of time
- Surfers also have the ability to “teleport”
  - They get bored
  - They reach a page with no links

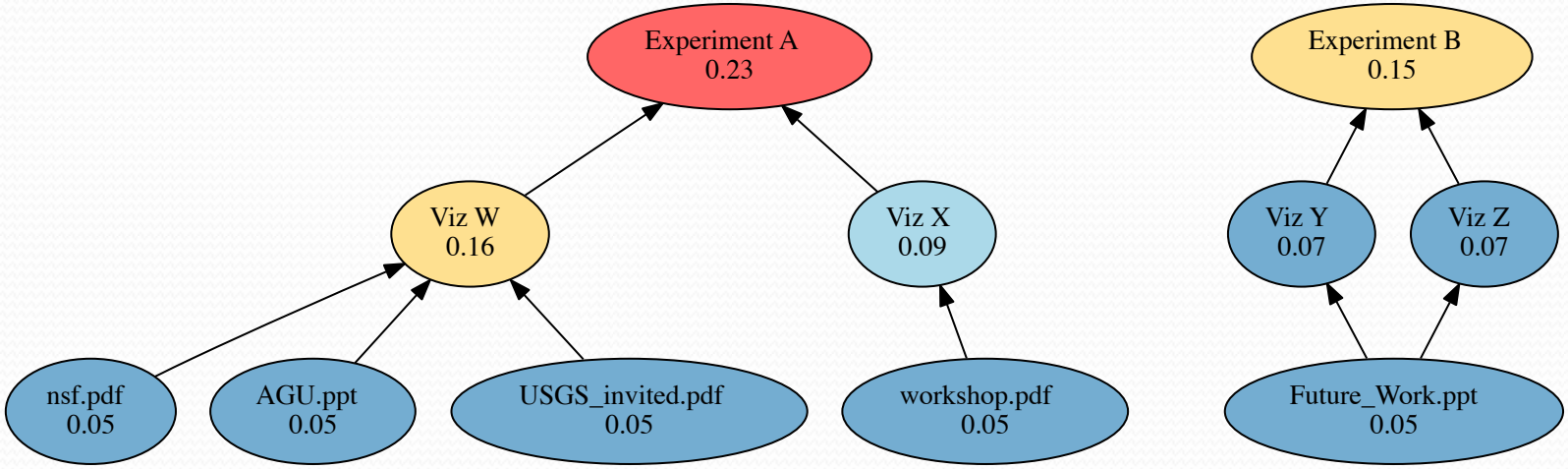
# Provenance Based Ranking

- An infinite number of scientists randomly follow inheritance links or open random files for an infinite amount of time.
- Provenance graph is a Directed Acyclic Graph, unlike the web, so teleport is more important
- To reduce focus on the oldest files, the random file open is weighted to favor newer files

# Provenance Graph Example

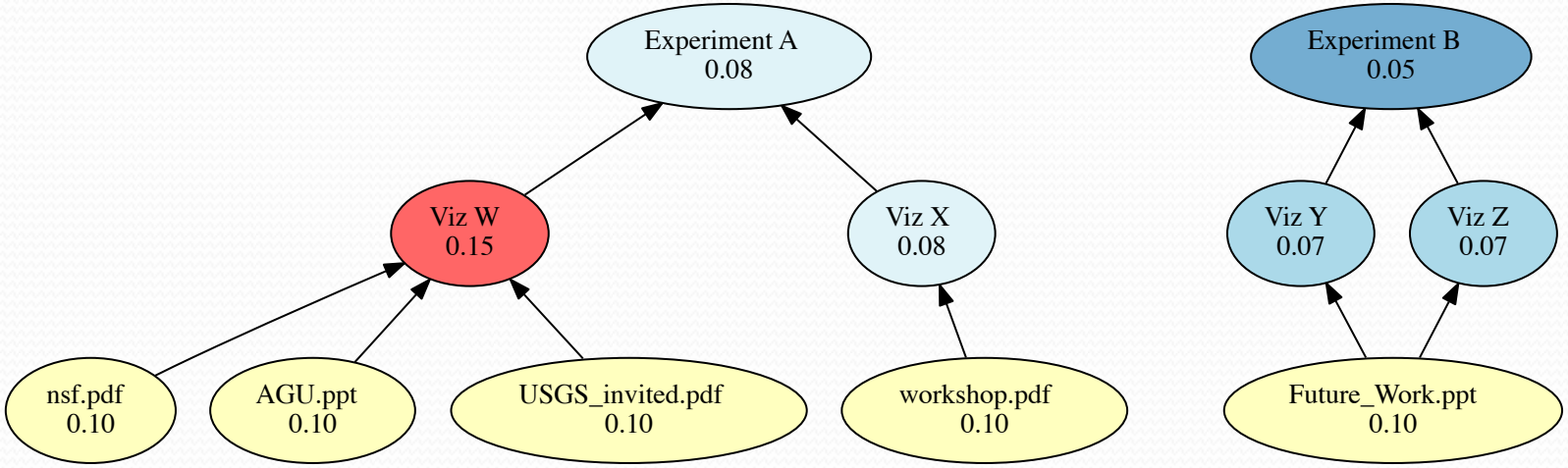


# PageRank Example




- The web has no roots per se
  - Provenance does
- PageRank focuses too much on them

# Provenance Rank example




- Use teleport function weighted by the height of the node
- Can focus on more intermediate files

# Conclusions

- Scientists need better tools to manage metadata and search their data

- Our system directly addresses the issues identified by the scientists we talked with
  - Provenance information creates correlations among files
    - Automatically relates similar files
  - Unified search space allows scientists to find files
    - Regardless of which storage system they are stored on
  - Ranking query results increases search speed
    - Identifies files important to the scientist

# Acknowledgements


- Scientists of LANL, PNNL and NOTUR for their time
- Meghan Wingate McClelland for organizing the LANL interviews
- John Johnson for organizing the PNNL interviews
- Department of Energy Office of Science
- Department of Energy's Petascale Data Storage Institute (PDSI)
- Center for Information Technology in the Interest of Society (CITRIS)
- NSF Center for Research in Intelligent Storage (CRIS)