

# sVC32 Collaboration and Virtualization using Active Cloud Engine

**Stephen Edel**

IBM SONAS Product Manager

IBM Storage

**Edge2012**

Expect More from Your Storage

June 4 – 8 Orlando, Florida



**This presentation explores how IBM Active Cloud Engine (ACE) in IBM Scale Out Network Attached Storage (SONAS) enables customers to collaborate and share information both within and across globally distributed locations. Discover how the future virtualization capabilities of ACE will enable migration and sharing of data between systems, to open up new opportunities to help customers manage their data. Real-world customer use cases will be discussed**

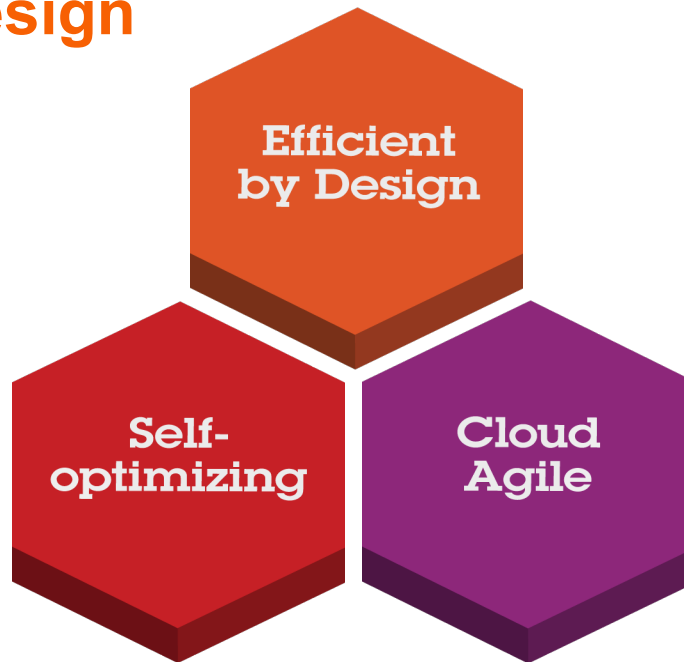
**With many thanks to John Sing for Contributions to this presentation**



# Introducing IBM Smarter Storage for Smarter Computing



**Efficient by Design**



**Self-optimizing**

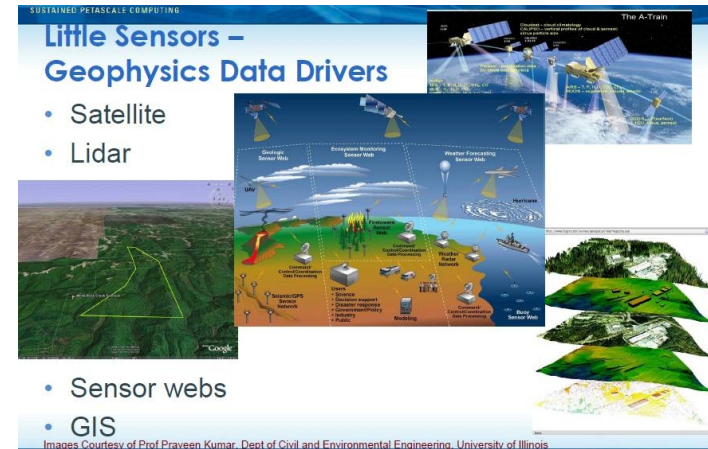
**Cloud Agile**



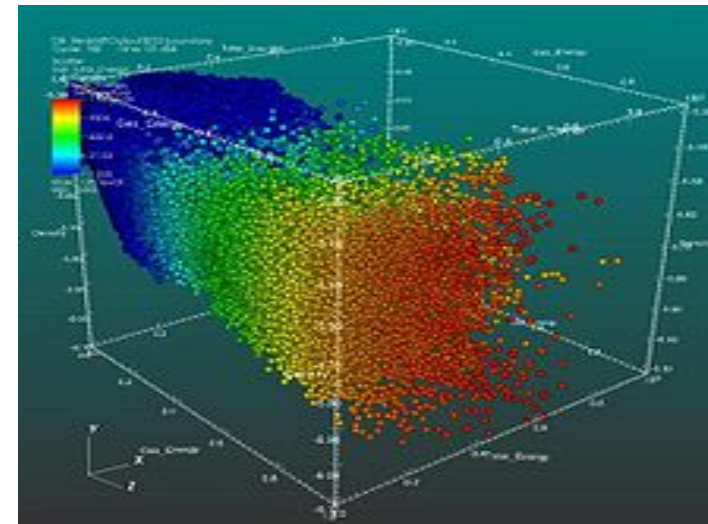
# Business Concerns of Unprecedented File System Data Growth



- **New and larger digital content types continuously evolving**
  - File sizes are growing to TBs
  - Moving from still images to HD video
  - Millions of files to Gazillions of objects
  - More kinds of files and objects
  - Volumes are reaching PBs per day
- **Greater collaboration requirements**
  - Creations points are many and everywhere
  - Collection points used to be few but now many, spread globally
- **New requirements to meet mission deployment / time to delivery**
  - Analysis needs information instantly
  - Access expected anywhere, anytime
  - Analysts no longer only in Washington, are deployed everywhere in the world
- **Mandates for longer retention periods and requirements for Archive & Compliance**



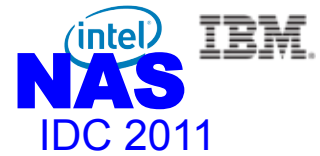
IEEE Massive File Storage presentation, author: Bill Kramer, NCSA:  
[http://storageconference.org/2010/Presentations/MSST/1\\_Kramer.pdf](http://storageconference.org/2010/Presentations/MSST/1_Kramer.pdf)



Source: Wikibon March 2011



# Fundamental, common elements to consider when evaluating a scale-out NAS solution:



- Multidimensional scalability: Scalability isn't just about expanding hardware capacity ... The best solutions will allow you to **scale each dimension** independently....
- Storage efficiency. Solutions that enable easy adoption of a continually expanding range of storage asset optimization technologies ... The best solutions will allow you to introduce or recalibrate these capabilities **without major disruptions or system migrations.**
- Intelligent Information Management. ... The best solutions will provide tight integration with **advanced metadata, index, and analytics solutions...** (that enable more intelligent information management as volumes of data sets grow and new use cases proliferate).



# Top requirements for Scale-Out NAS solution

ESG analyst: Terri McClure  
December 2011



1. Global namespace
2. Clustered, managed as a single entity
3. Ability to scale bandwidth, processors, and storage independently
4. Power-efficient
5. Self-managing
6. Self-healing
7. Transparent data mobility
8. Tiered storage support
9. Feature/functions :

If going after enterprise IT then snapshots/clones/remote mirrors

If going after archive then deduplication/compression



# Gartner's recommendations to Scale Out NAS vendors



Companies selling high-end scale-out NAS should:

- Continue investing in building the ecosystem and complete solutions for managing large amounts of unstructured data
- Invest in R&D to develop practical and cost-effective backup and recovery solutions
- Enhance global access capabilities to address the increasing need for global sharing and collaboration, as well as cloud initiatives
- Publish SPECsfs Network File System (NFS) and Common Internet File System (CIFS) benchmark results based on real-world configurations and workload types to make the benchmark more meaningful
- Develop more-affordable solutions for long-term retention of unstructured data

Gartner analyst: Pushan Rinnen  
November 2011



# Why IBM ACE is important to your clients – and You !

*Let's Take a Look at the Recent Evolution of Storage Marketplace*

Mid-1990's – **Data on IP Network = NAS (Network Attached Storage)**

Early 2000's – **Block Level Virtualization = Storage Virtualization**

Mid/Late- 2000's – **Global Namespace = SONAS (Scale Out NAS)**

2011+ – **File Virtualization / Cloud Storage = Active Cloud Engine**





- **What is IBM Active Cloud Engine?**
  - **Policy-driven engine** that helps improve storage efficiency by automatically
    - Distributing files, images, and application updates to multiple locations \*
    - Identifying files for backup or replication to a DR location
    - Moving desired files to the right tier of storage including tape in a TSM hierarchy
    - Deleting expired or unwanted files
  - High-performance: can scan billions of files in minutes with scale out nodes
- **What client value does Active Cloud Engine deliver?**
  - Enables ubiquitous **access to files** from **across the globe** \*
  - Reduces networks costs and helps improve application performance by **distributing files closer to users** \*
  - Improves data protection by identifying candidates for backup or DR
  - Lowers storage cost by **moving files transparently** to the most appropriate tier of storage
  - Controls storage growth by transparently **moving older files to tape** and deleting unwanted or expired files
  - Enhances administrator productivity by **automating file management**



# IBM Active Cloud Engine Local Management Capabilities



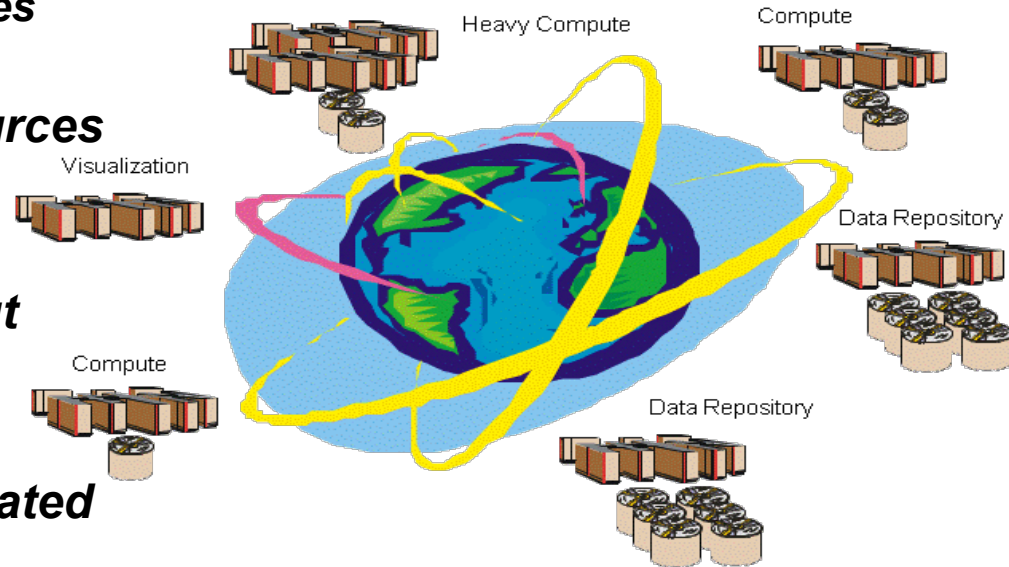
- Active Cloud Engine enables two types of policies
  - *File placement policies*
    - Used to automatically place newly created files in a specific storage pool
    - Files in the same fileset may reside on different storage pools
  - *File management policies*
    - Transparently move files to another internal or external storage pool
    - Delete, backup, incremental backup, restore, replicate
    - Remote caching, intelligent prefetching of subsets of files
    - Virtualize to single global namespace.



# IBM Active Cloud Engine Global Capabilities

Manage file system data with one unified management system for global file access through policies and automation while managing its lifecycle with reduced risk, complexities and cost

- **Break the dependence of an object from its location**
  - **Files (or objects) move around the sites**
- **Allows sites to contribute their resources to the global file system**
- **Allows a single gold copy of data, but copies also at the right place**
- **Allows data to be permanently replicated for fault-tolerance and performance**
- **Supports disconnected operations for network deprived locales**



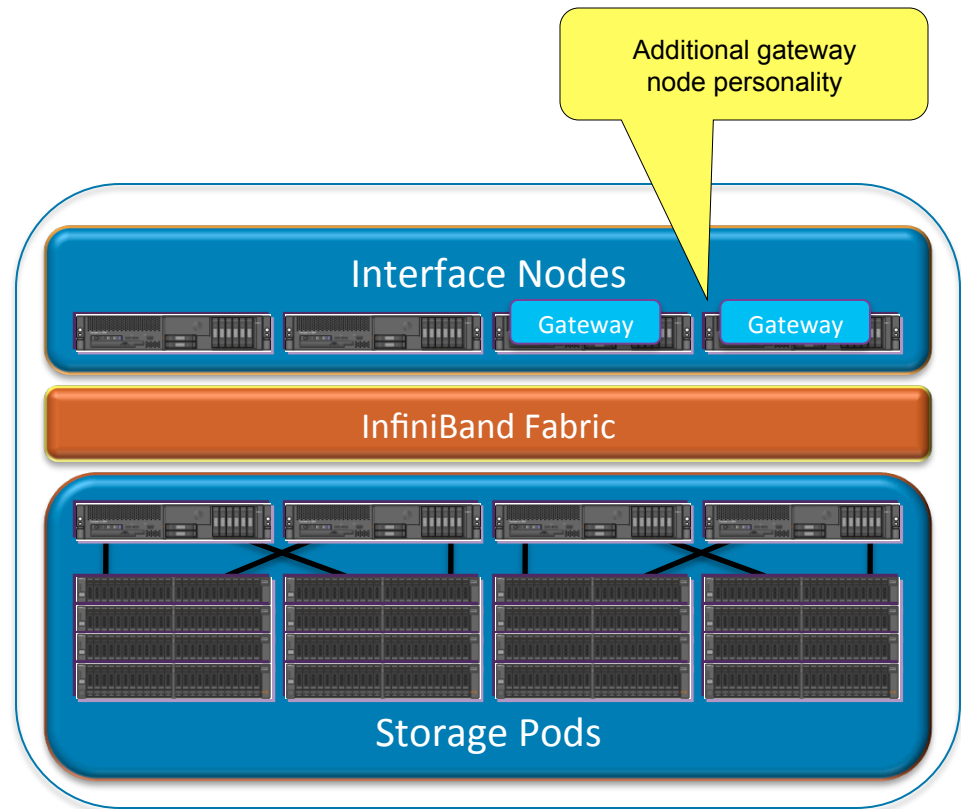
# IBM Active Cloud Engine - Cache Cluster

## Data Caching

- Data requests are handled locally if the requested data is cached
- Requests for data not in cache are forwarded to the gateway nodes to retrieve the data from the home cluster.

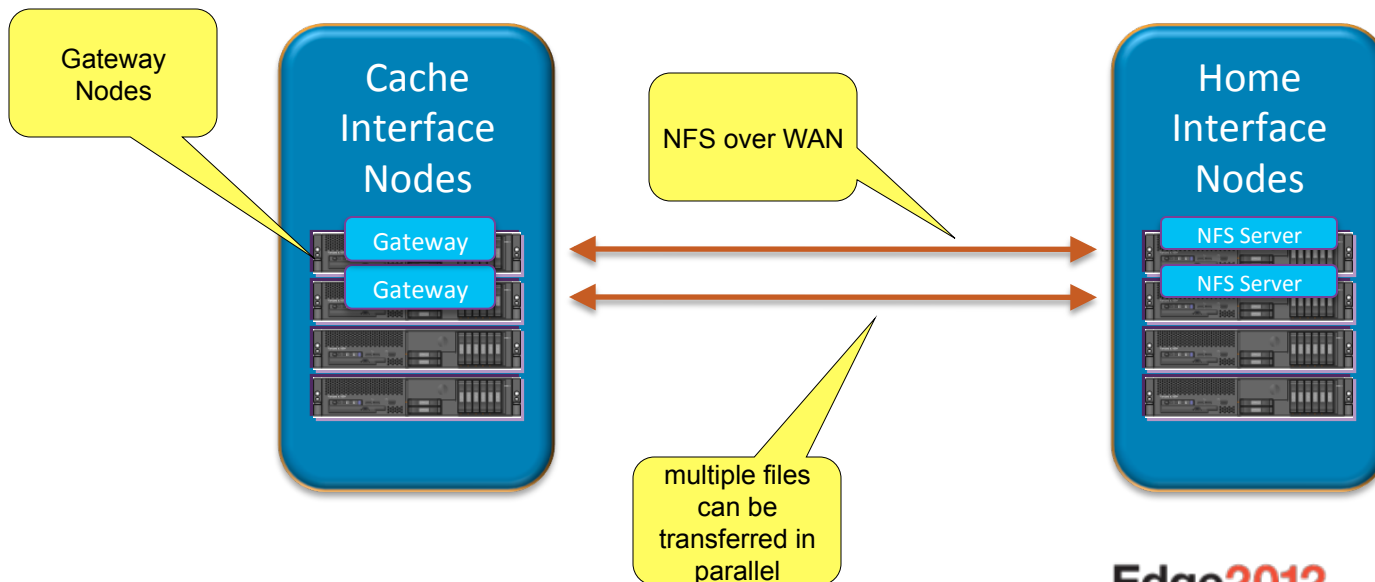
## Gateway Nodes

- IBM SONAS interface nodes take on a personality known as a **“Gateway” node**
- Gateway nodes communicate with the home cluster
- Any interface node can function as a gateway node
  - And still function as an interface node



# IBM Active Cloud Engine – Home Cluster

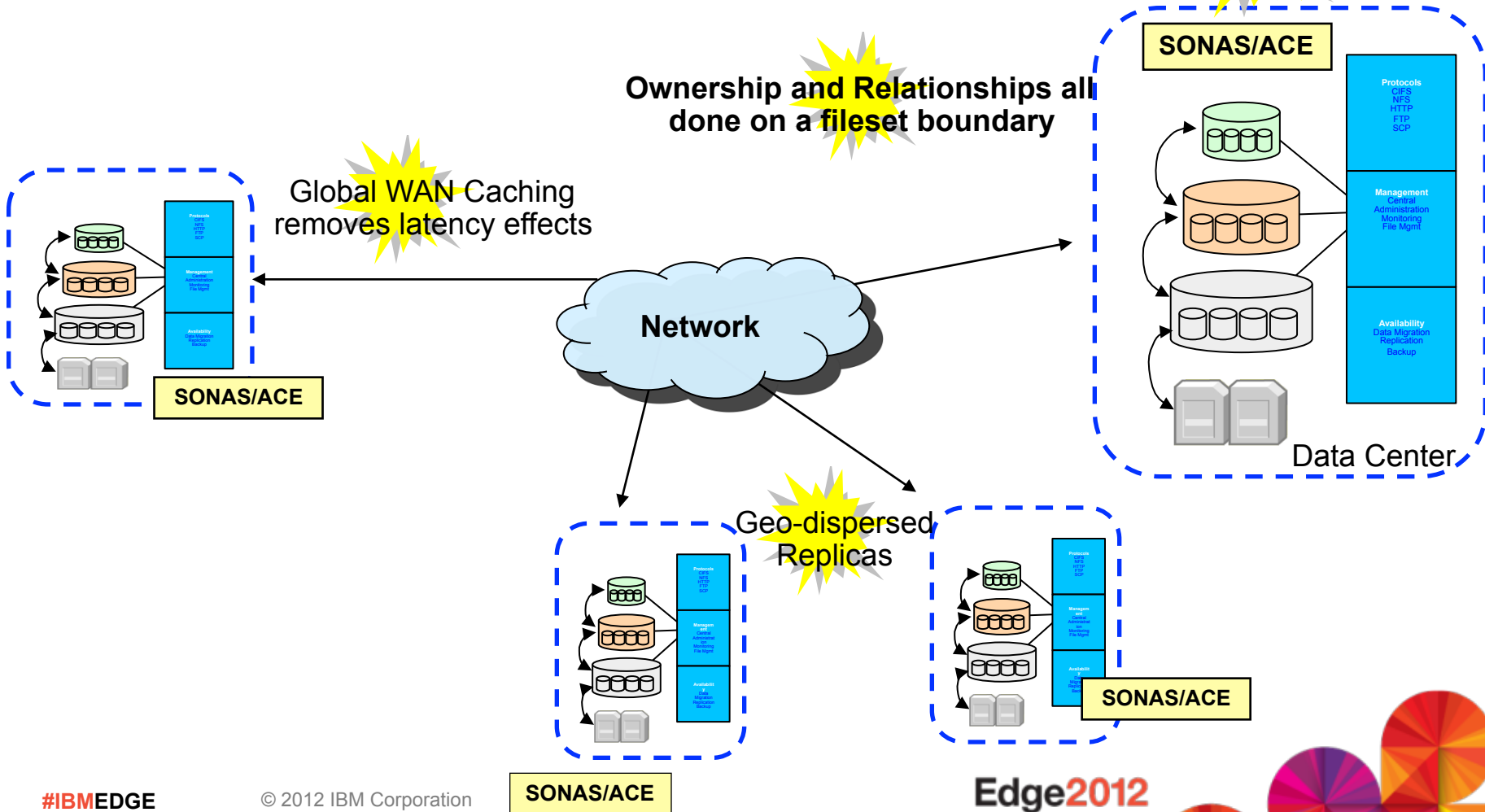
- When data is not in the cache, the Gateway nodes request data from the home cluster
- Permissions are checked before the file is copied from the home cluster
- The Home Cluster exports data using the NFS protocol using one or more interface nodes



# Active Cloud Engine Features and Advantages

**Extends “common” namespace & ILM/HSM Policy engine to global geo-dispersed file storage**

“Global Namespace, not just “Common” Namespace

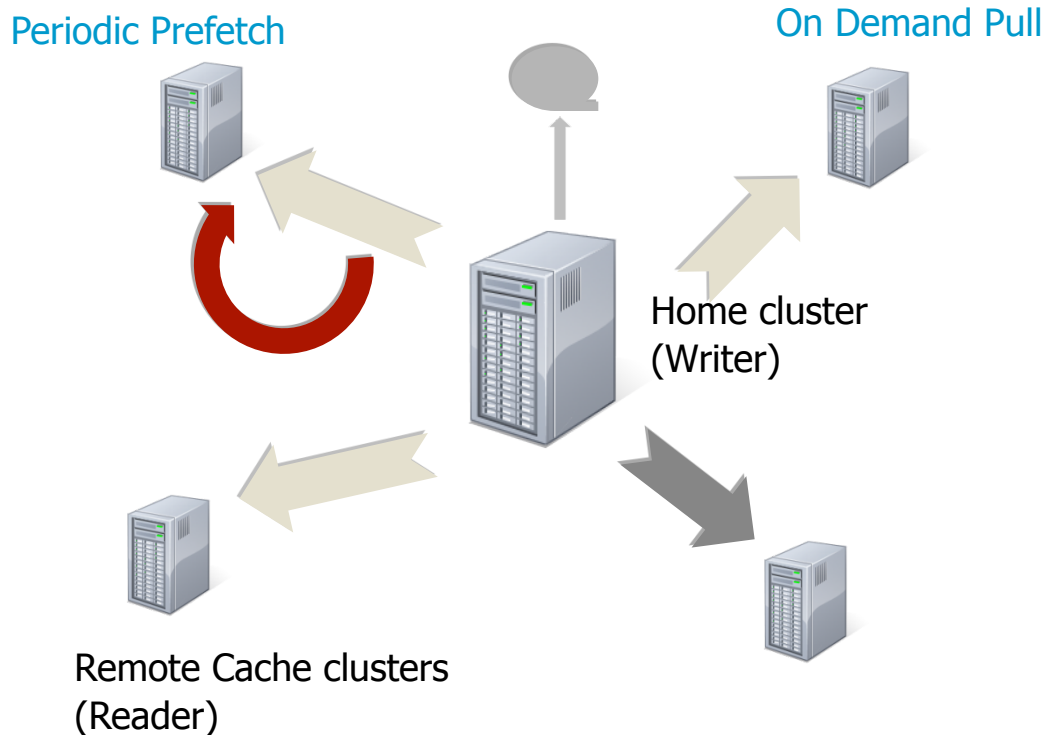


# Active Cloud Engine Modes (Fileset level)

- **Read Only Cache**
  - Cache can only read data, no data changes/updates allowed for that fileset at that cache location
- **Exclusive / Single Writer Cache**
  - Cache can write data to each fileset it owns. Home cluster can't change as it does not own the fileset. Other peer caches have to be setup as read only caches.
- **Local Update**
  - Data is cached from home and changes are allowed like Single Writer mode, but changes are not pushed to home.
  - Once data is changed the relationship is broken i.e cache and home are no longer in sync for that file
- **Change of Modes**
  - SW & RO mode caches can be changed to any other mode
    - Requires cli to be run on cache & home
  - LU cache can't be changed



# Use Case: Remote Cache Sites operating in Read Only Mode



- Central Data Center is where data is created, maintained, updated / changed.
- Branch locations can periodically prefetch (via policy) or pull on demand files as requested
- Data is revalidated back to Home cluster when accessed by a user at the cache site
- Customers Use Cases:
  - Financial branch offices
  - Dispersed Government locations
  - Media Distribution Centers
  - WW Research Labs
  - Regional Hospitals



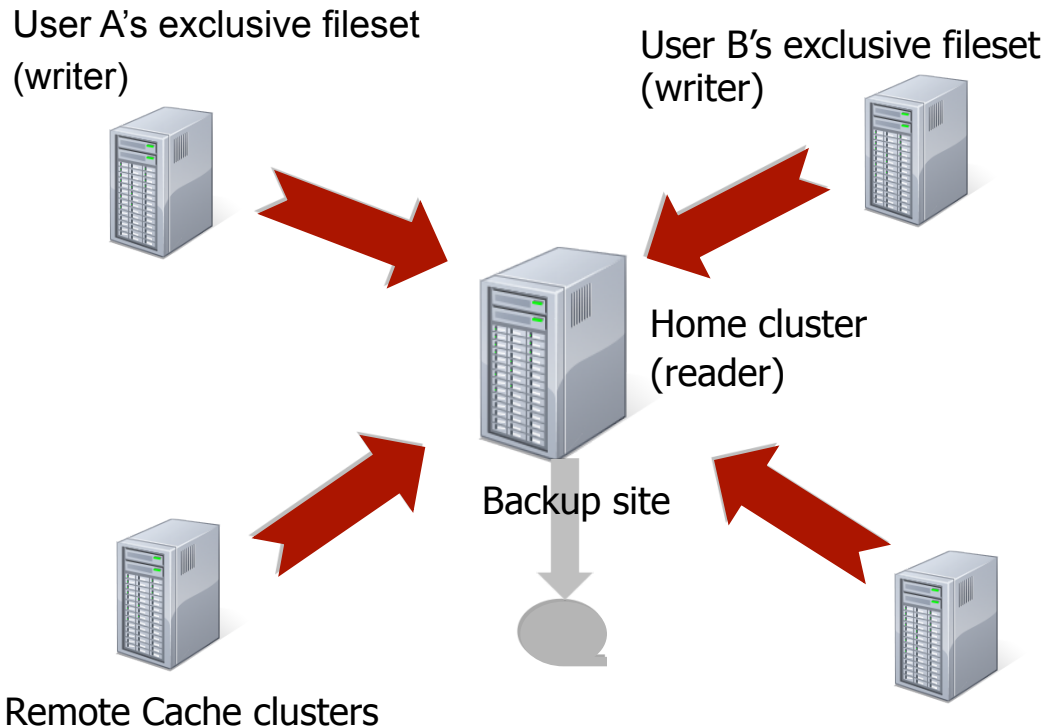


# Read Only Scenarios (cache validate/miss)

- If data is not in cache, pull file attributes and create on demand (lookup, open, ...)
- **Directory Traversal**
  - On first access, entire directory is read remotely (REaddir) and directory entries stored locally
- **On a later data read**
  - Whole file is fetched over NFS and written locally
  - Multiple files can be read in parallel across multiple nodes
  - Application can continue after required data is in cache while the remaining file is being fetched
  - In addition, a cache-site can be pre-populated before any access
- **On a cache hit**
  - Attributes are *revalidated* depend on revalidation delay
    - (done periodically on lookup and open)
  - If data hasn't changed it is read locally
- **On a disconnected mode access**
  - Consider cache valid and allow open without revalidation
  - Files not cached are returned as not existing



# Use Case: Remote Cache Sites having Single Writer Ownership



- Each site has defined filesets which own the data and serve as their local cluster.
- Remote sites push data back to central Data Center for additional DR processes and for access by others

## Use Cases

- Federated Institutes of Research with a common repository for access
- Data Repositories or Archives
- Oil/Gas exploration delivering data to a Central Site
- Universities sharing a dedicated home file system for students but also linking other campuses
- Telco headquarters tracking logs/records
- Genomic sequencing results stored in a Common location



# Single Writer Scenarios

## Asynchronous Updates (write, create, remove)

- **Updates at cache site are pushed back lazily**
  - Masks the latency of the WAN
  - Data is written to SONAS at cache site synchronously
    - GW node queues the update for later execution
  - Performance identical to a local file system update
- **Write back is asynchronous**
  - Configurable async delay
  - GW nodes queue updates and write back to home as network bandwidth permits
  - Write back can coalesce updates and accommodate out-of-order and parallel writes to files and directories, thus maximizing WAN bandwidth utilization
- **Users can force a sync if needed**



- **Staleness Control**
  - Defined based on time since disconnection occurred
  - Once cache is expired, no access is allowed to cache
  - Manual expire/unexpire option for admin
  - Allowed only for RO mode cache
  - Disabled for SW & LU modes as they are the recognized sources of data
- **Disconnected**
  - No connectivity with home
  - Updates are queued
  - Reads are serviced locally
  - All access stops after expiration
- **Resync Needed**
  - An error or corruption detected at home
  - ACE will try to fix some errors
  - Need to retry if error is temporary
  - Needs admin intervention for data corruption error



# Cache Eviction Scenarios

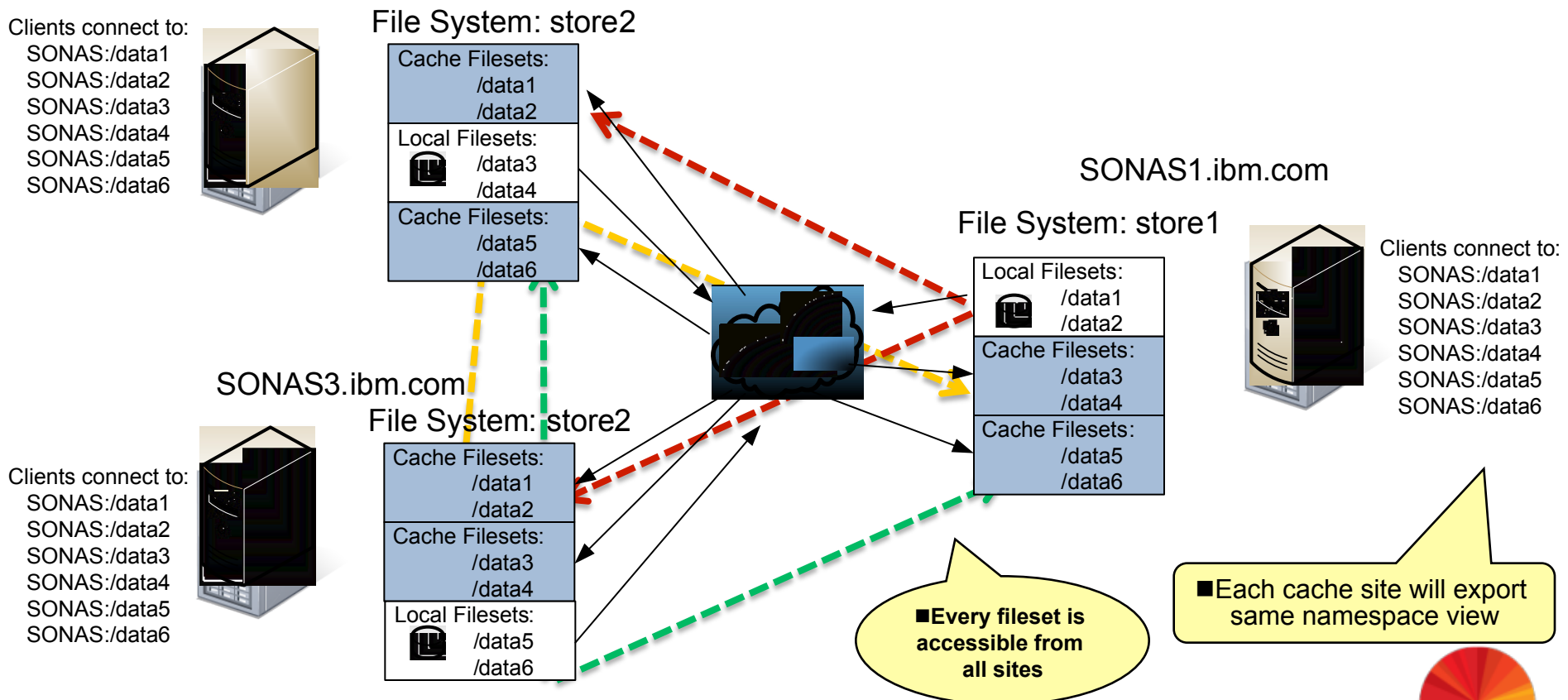
- When is it needed?
  - When cache cluster is smaller than home cluster
  - If data fills up in cache faster than it can push to home.
  - Need to create space for caching other files or space for incoming writes.
  - Eviction is linked with fileset quotas.
- Cache eviction is triggered automatically
  - fileset usage level goes above fileset soft quota limits.
  - Eviction tries to bring down usage below 5-10% of quota limits by de-allocating blocks for files in cache.
  - The dirty files with pending flushes are not looked at by eviction.
- It can be triggered manually also if required



# Use Case: Global Namespace (Mesh)

- ACE can be used as a mesh...that is any fileset in any location can be a cache or a single writer fileset to any other location
- Up to 3000 independent filesets per file system

SONAS2.ibm.com



# IBM SONAS with Active Cloud Engine: Example use cases

## Surveillance and Monitoring

High resolution image capture, data analysis, and results distribution (e.g. Weather)

## Digital Media

High performance, simplified management alternative for widely varying use cases in digital media environments.

## University Collaboration

Link Institutes and Researchers for Data Sharing and analysis

## Energy & Geo-Sciences

Energy exploration and geo-sciences require huge addressable namespaces for analytics and very high performance.

## CAE

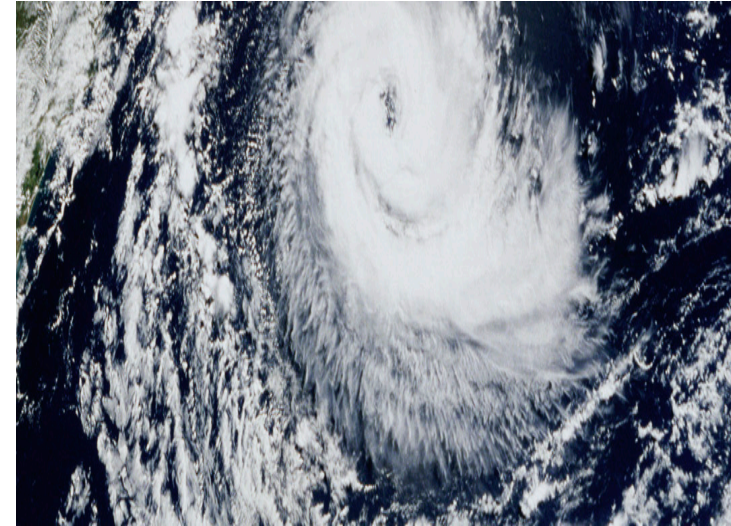
Auto / Aero / Electronics design processes experiencing rapid file-centric storage growth as simulation expands.



# Real Time Image Data Capture/ Surveillance and Monitoring

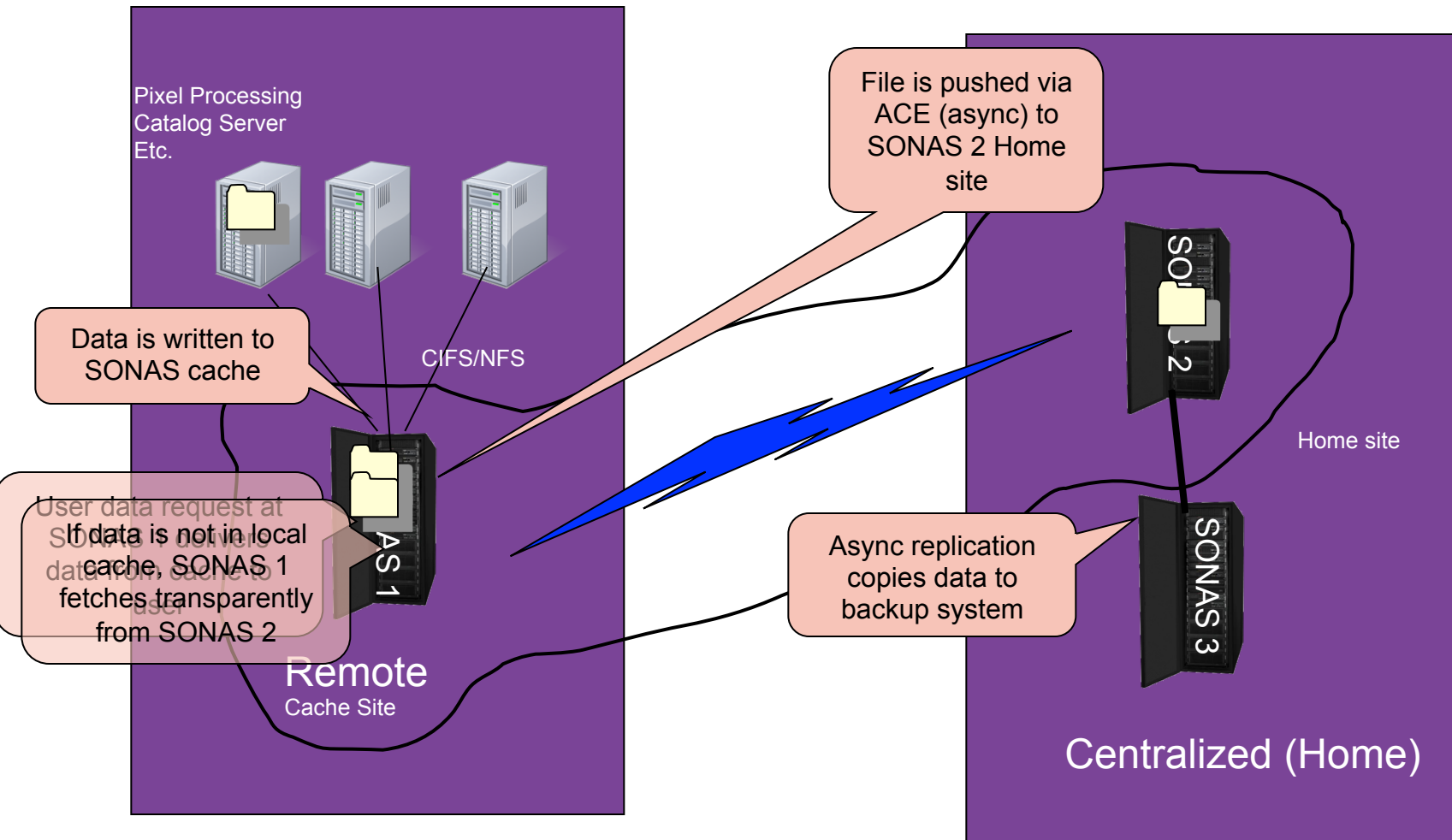


- High Resolution weather images are recorded from planes or satellites
- Information is directly loaded into SONAS systems at remote locations
- IBM Active Cloud Engine allows transfer of these images to a Central Repository
- Weather experts at other sites can then access the data from the Central Repository, perform analysis and provide guidance and information to be disseminated to others





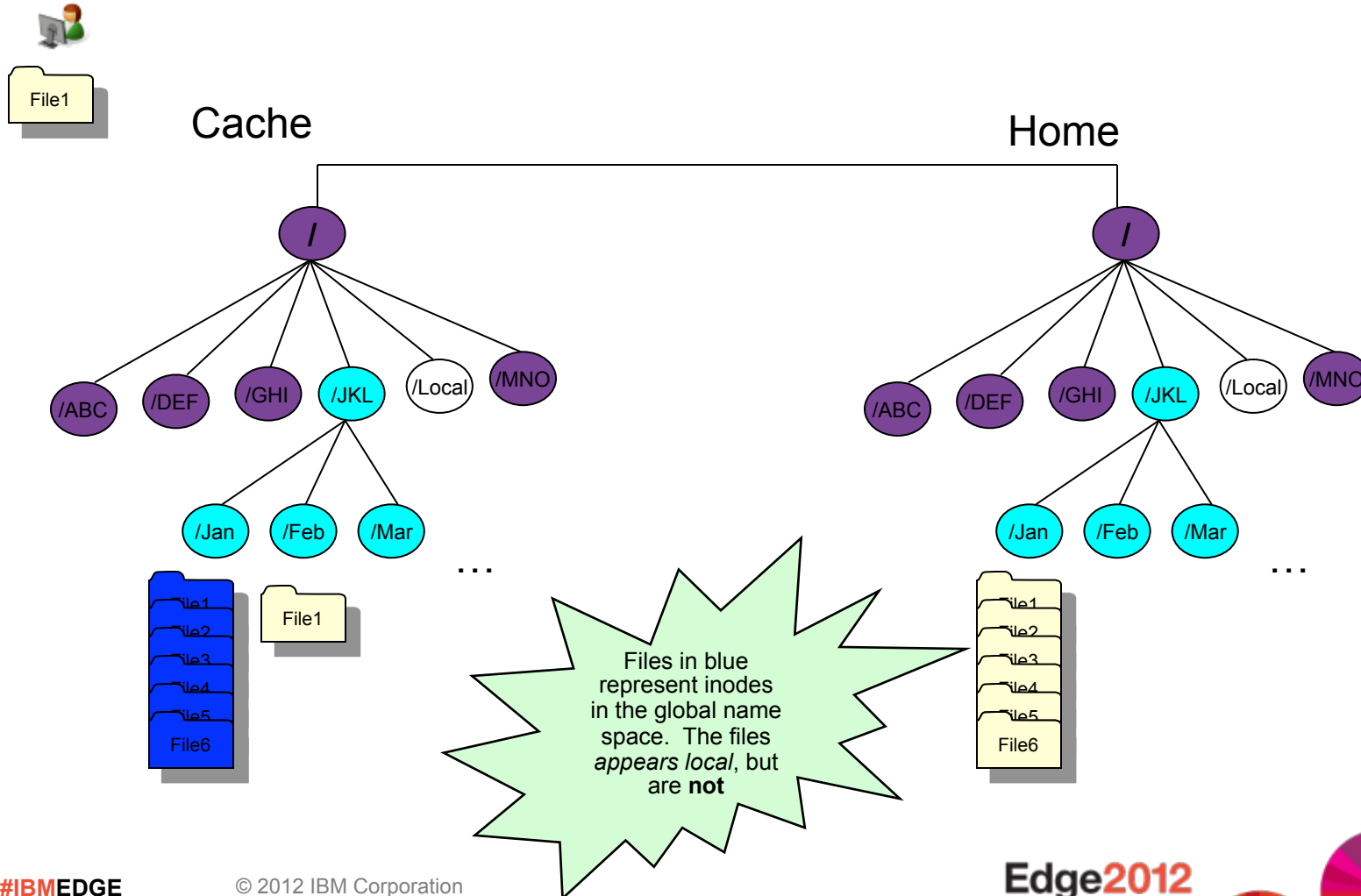
# Example Applying ACE caching technology to Surveillance / Monitoring environment



# Making the Cloud *Active* - Write a File



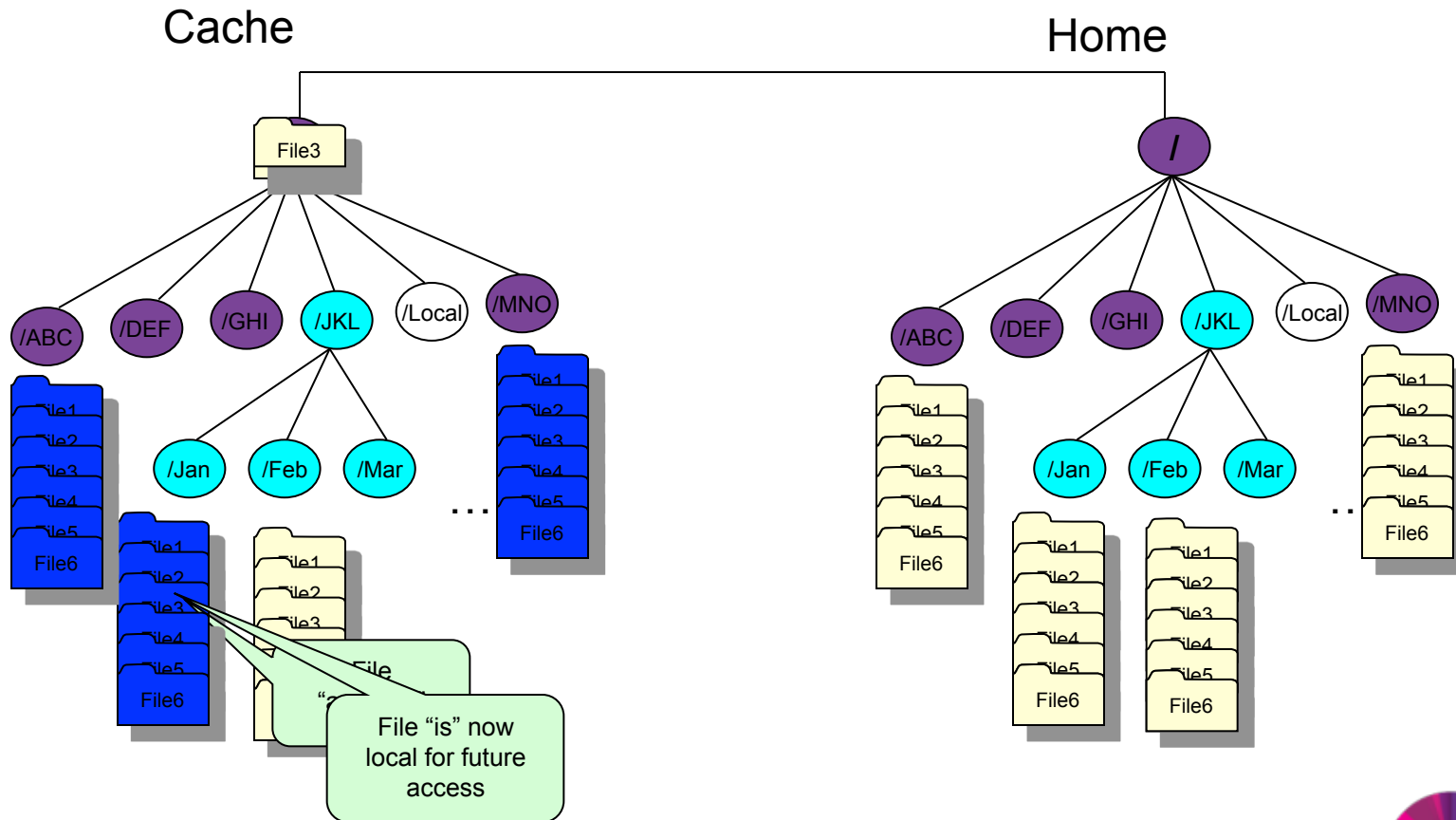
Cache cluster user writes a file on February 1<sup>st</sup> that is copied immediately or via a schedule to Home Site



# Read a File from

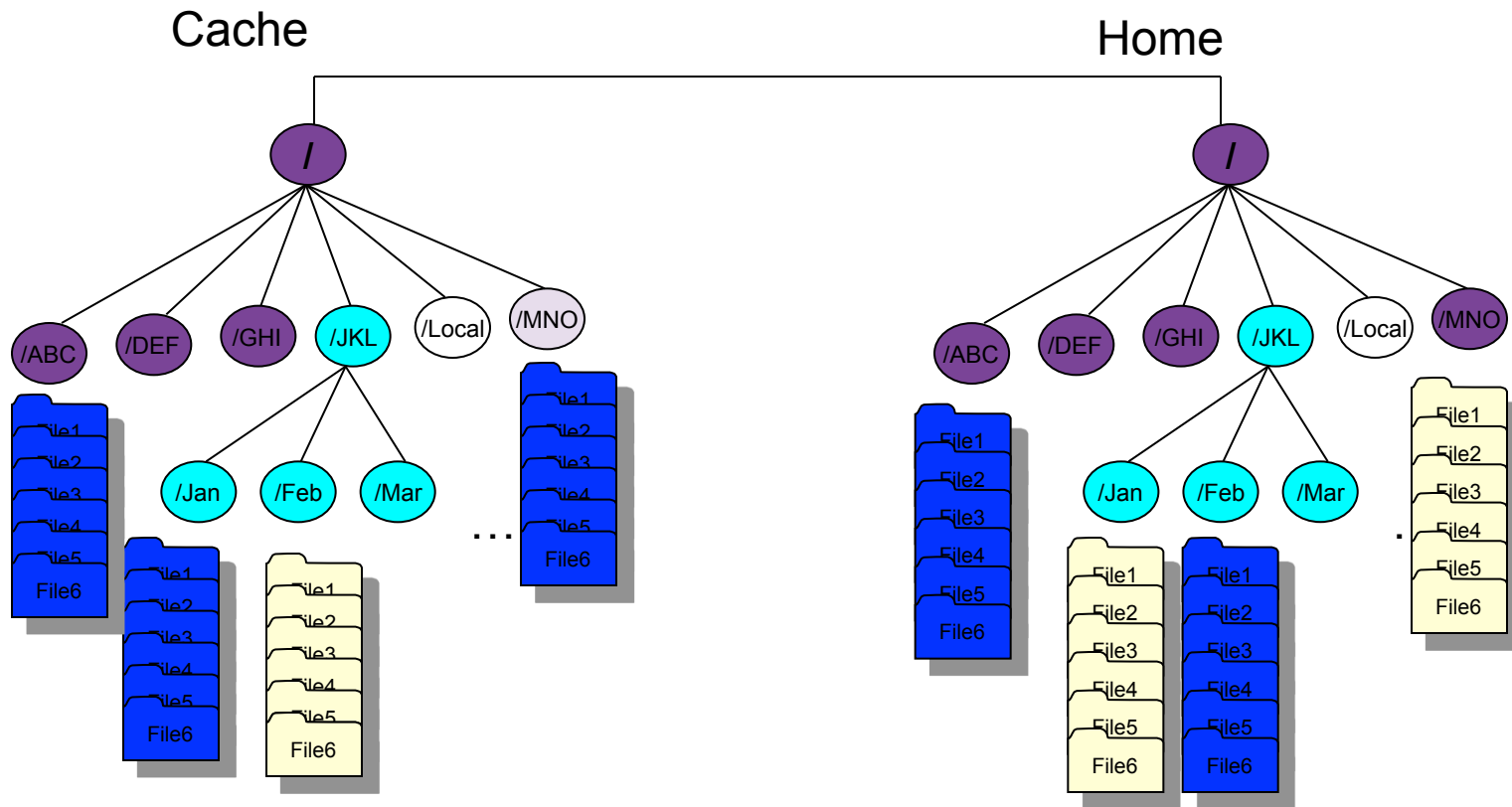
# “Home”

Cache user “double clicks” on /JKL/Jan/file3 in his directory that appears local but is actually at the Home site



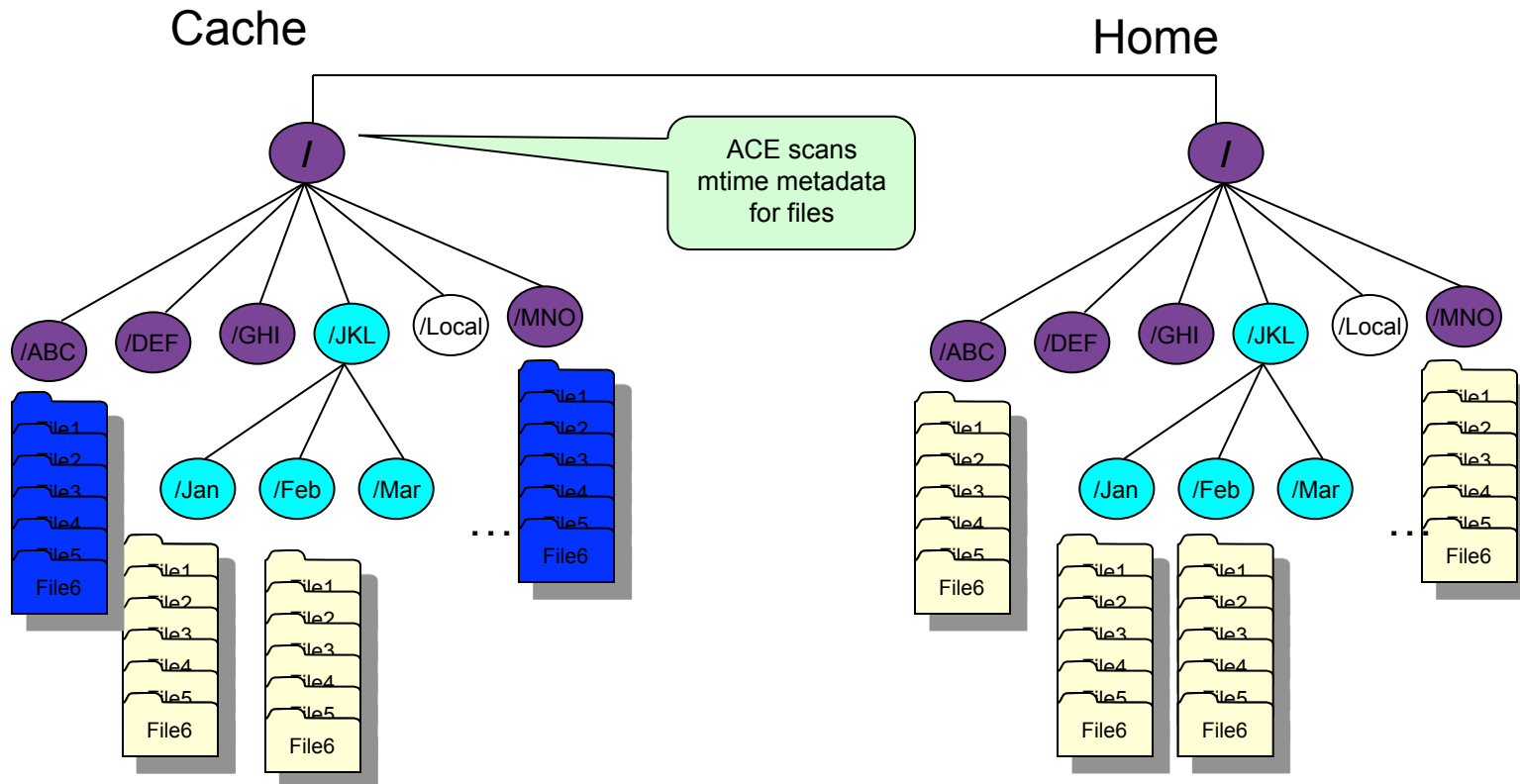
# Cache Cluster Pre-population

Policy change to pre-fill the Cache with new version of data for “MNO”



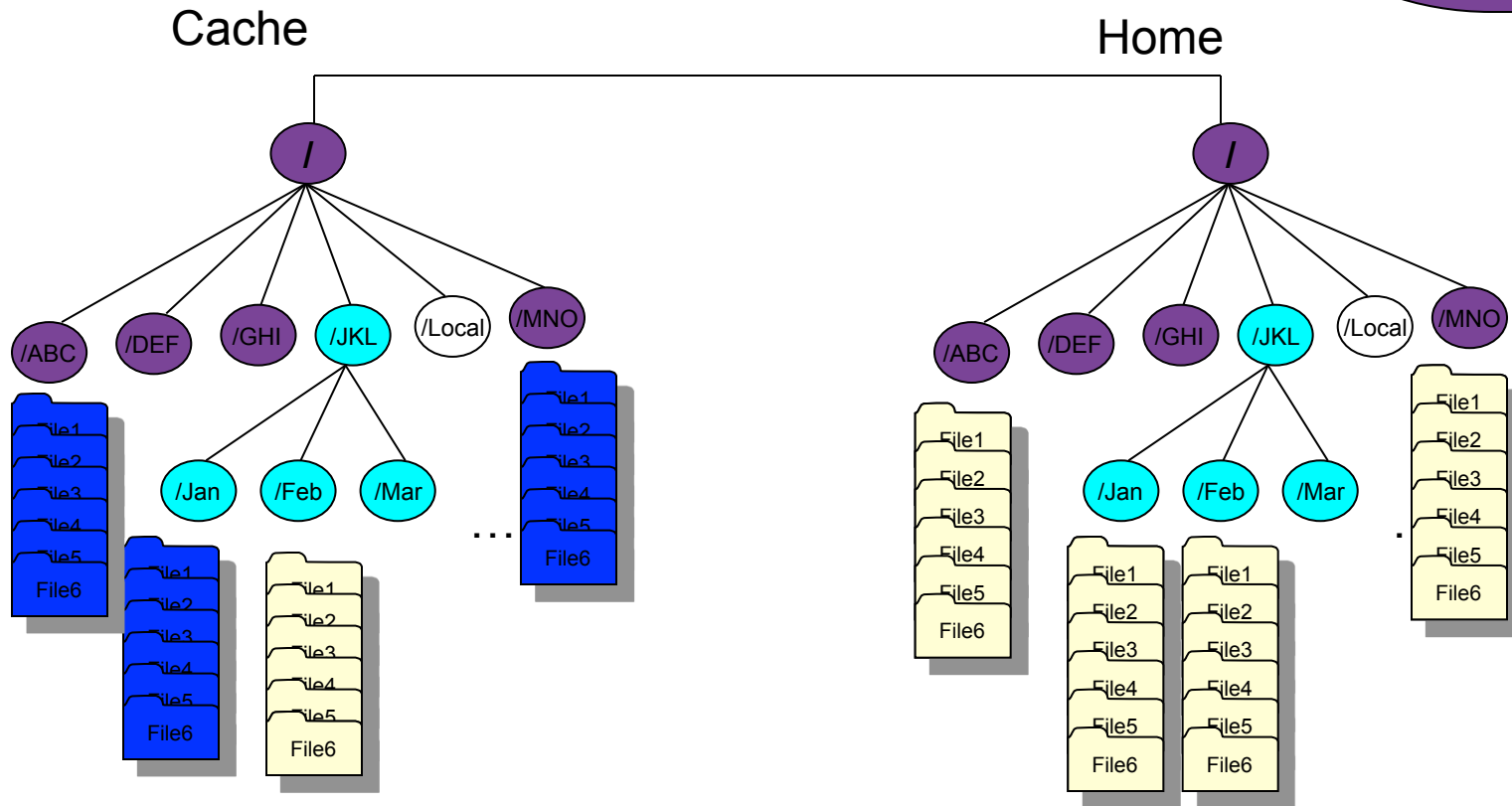
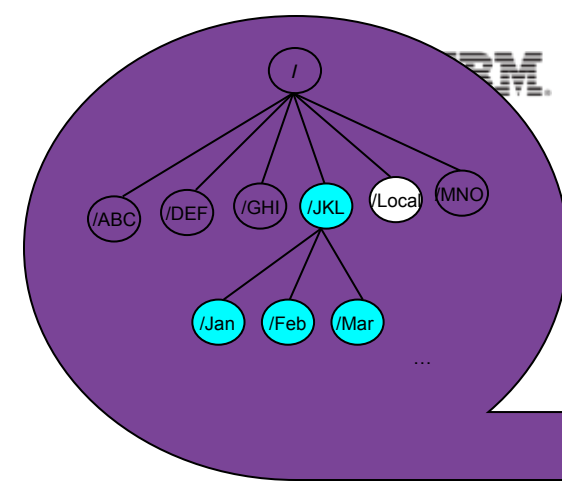
# Clearing the Cache on policy

Clear the Cache of all files not modified in the last 30 days



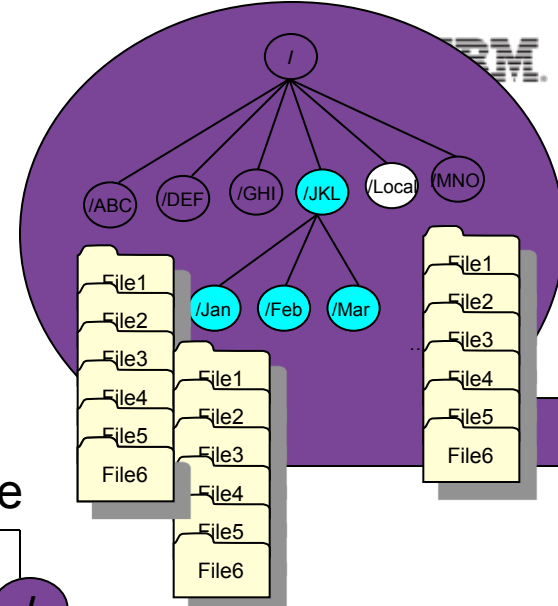
# Archive to tape

Archive data 30 days old to tape by policy



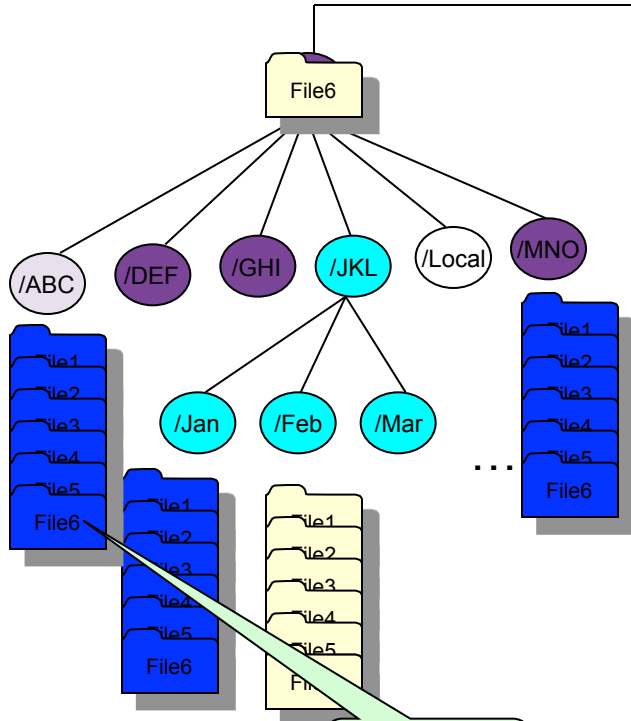
# Demand File Pull

Cache User opens up /ABC/File6

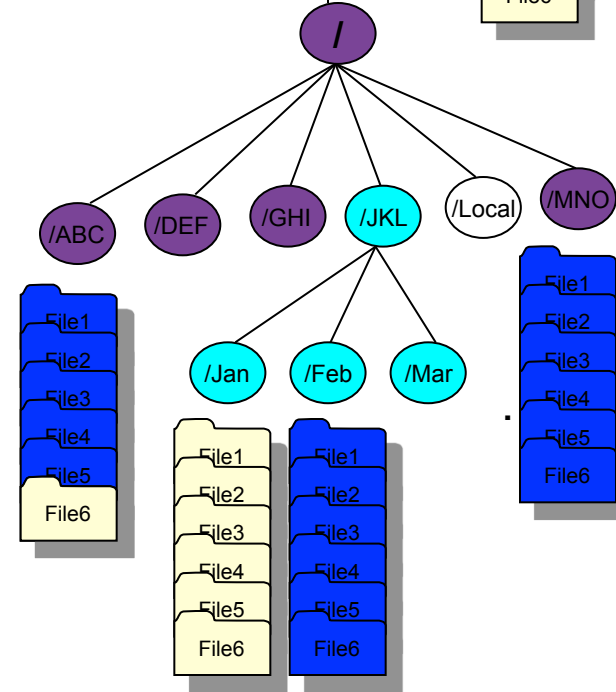


Cache

Home



File 6  
"appears"  
local



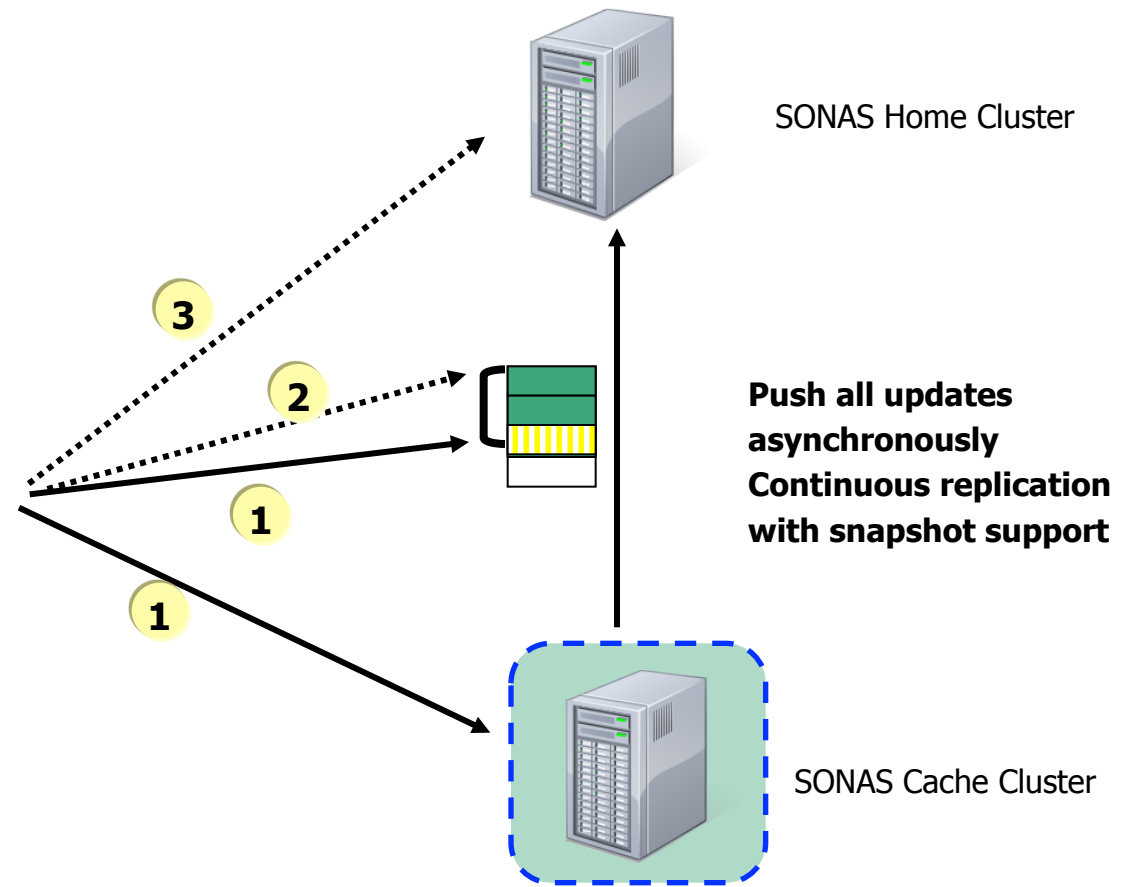
# Peer Snapshot Consistent Replication

**1** Take a fileset snapshot at cache cluster site.  
Mark the point in time in the “write-back queue”

**2** Push all updates up to point in time marker

**3** Take a snapshot of fileset at the home  
Update management tool state of last snapshot time and ids

Multi site snapshot scripts





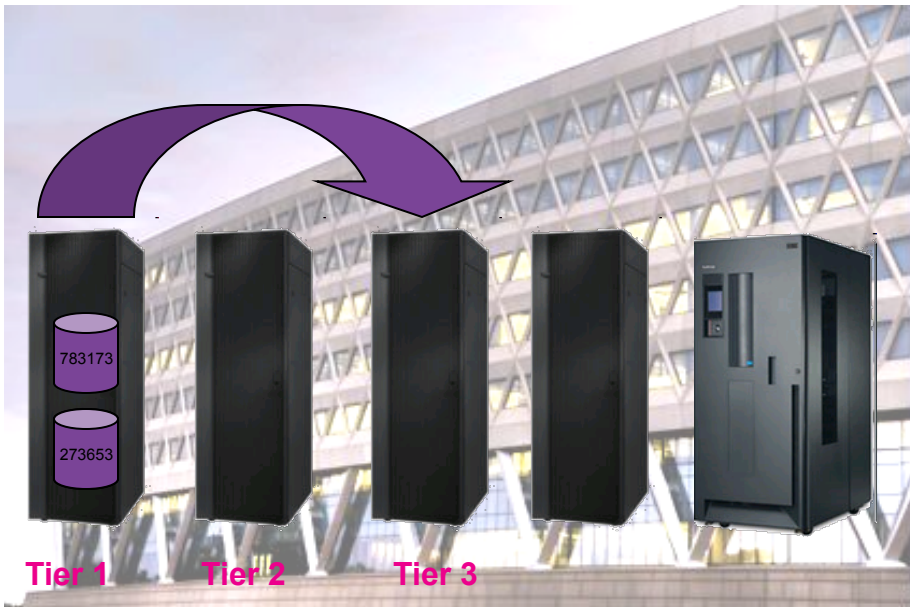
# IBM SONAS Global Policy Management



Filename	type	Size	Owner	Creation Timestamp	Last Access Timestamp	Last modified Timestamp	Permissions	World Coordinates	Acquisition time	Country code
783173	PNG	10M	User1	Apr 1, 2009	Jun 1, 2009	Apr 1, 2009	RXX	Latitude/ Longitude	2009-3-30 T 10:45	Brazil
273653	RAW	100G	User9	Jan 2, 2011	Feb 6, 2011	Jan 15, 2011	RWX	Latitude/ Longitude	2010-10-30 T 10:45	Canada

Find all the files of “filetype” PNG where “creation timestamp” > than 1 year old and “last access” > 6 months” and move them to tier 3

Specific Metadata

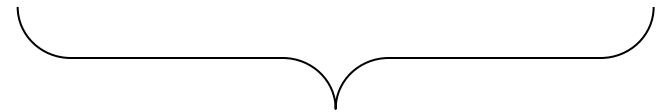




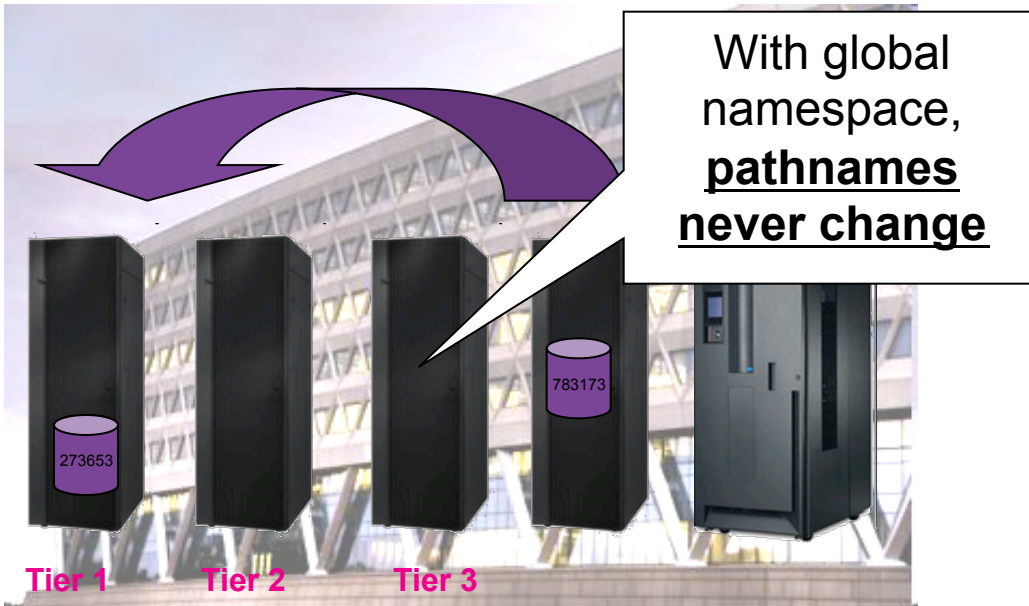
# IBM SONAS Application Specific Storage

Filename	type	Size	Owner	Creation Timestamp	Last Access Timestamp	Last modified Timestamp	Permissions	World Coordinates	Acquisition time	Country code
783173	PNG	10M	User1	Apr 1, 2009	Jun 1, 2009	Apr 1, 2009	RXX	Latitude/ Longitude	2009-3-30 T 10:45	Brazil
273653	RAW	100G	User9	Jan 2, 2011	Feb 6, 2011	Jan 15, 2011	RWX	Latitude/ Longitude	2010-10-30 T 10:45	Canada

Find all the files where country code="Canada" and move to tier 1



Specific Metadata



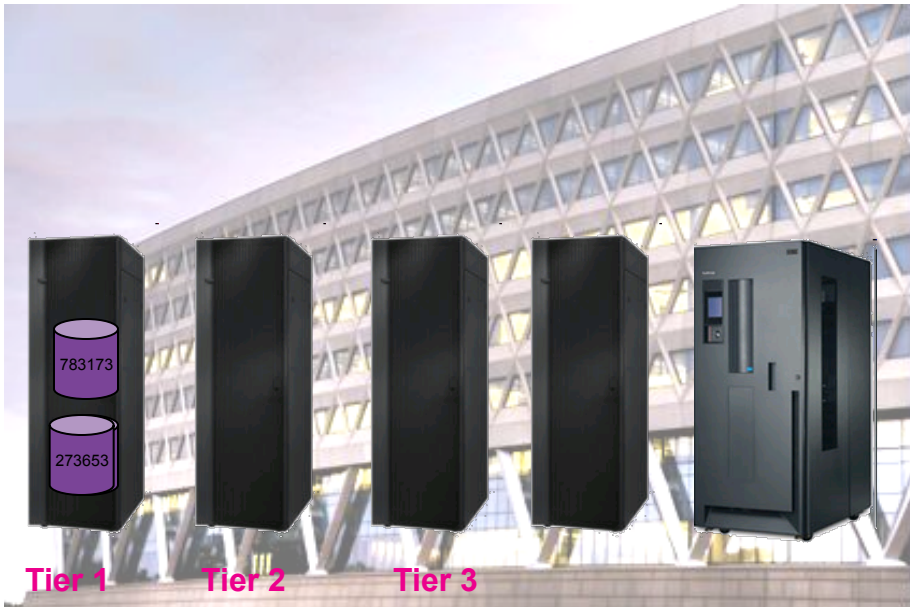
# Global Policy Management using IBM Active Cloud Engine



Filename	type	Size	Owner	Creation Timestamp	Last Access Timestamp	Last modified Timestamp	Permissions	World Coordinates	Acquisition time	Country code
783173	PNG	10M	User1	Apr 1, 2009	Jun 1, 2009	Apr 1, 2009	RXX	Latitude/ Longitude	2009-3-30 T 10:45	Brazil
273653	RAW	100G	User9	Jan 2, 2011	Feb 6, 2011	Jan 15, 2011	RWX	Latitude/ Longitude	2010-10-30 T 10:45	Canada

Replicate all the files where “Country Code”= Canada where “creation timestamp” < than 1 year old and push to “Brazil”

Specific Metadata



# State Government Consolidation

- Customer looking to link all key agencies across the state to capture key data at each city
- Goal is to consolidate local servers in each city.
- Lots of office documents in each agency and city



State Government

## Solution

- Solution consists of 2 SONAS gateways with V7000 storage located at the State Capitol
- ACE will be used in an exclusive writer mode to move files to the Capital
- Eliminates repetitive “auto-save” operations creating delays since ACE provides local access speeds
- Eventually, V7000 Unified will be the future remote site deployments at key cities for cost optimization



# Financial – Workload Optimization



Financial Institution

- Interested in building a file-based Content Distributed Network to enhance workload distribution and placement capabilities
- Ability to distribute workload in a data center today and to various compute farms - move data to applications
- Meta data (date, access time, permissions, and support of policies for data placement and data migration and replication.
- Central point of management, but with multiple copies in different locations
- Data resides at a central Data Center but seamlessly distributes data to edge cache sites
- Supports copying of data at remote sites back to home cluster

## Solution

- Solution consists of 2 SONAS gateways currently in an ACE relationship
- Testing underway demonstrating how ACE provides the key capabilities listed above



# Statements of Direction\* for IBM Smarter Storage - An integral part of IBM Smarter Computing

- IBM Active Cloud Engine™ will provide NAS Virtualization capabilities to enable transparent file migrations from selected NAS systems into SONAS or Storwize V7000 Unified.
- IBM Active Cloud Engine will allow users at geographically dispersed locations to both share and modify the same file (SONAS/ Storwize V7000 Unified)
- Using cloud enhancements and self-service cloud portal for SONAS and Storwize V7000 Unified, enterprises can implement a private cloud storage service where users, with a few clicks, can request & receive storage capacity, share files with other users, and administrators can easily monitor & report usage.

\*Statements of IBM's future plans and direction are provided for informational purposes only. Plans and directions are subject to change without notice



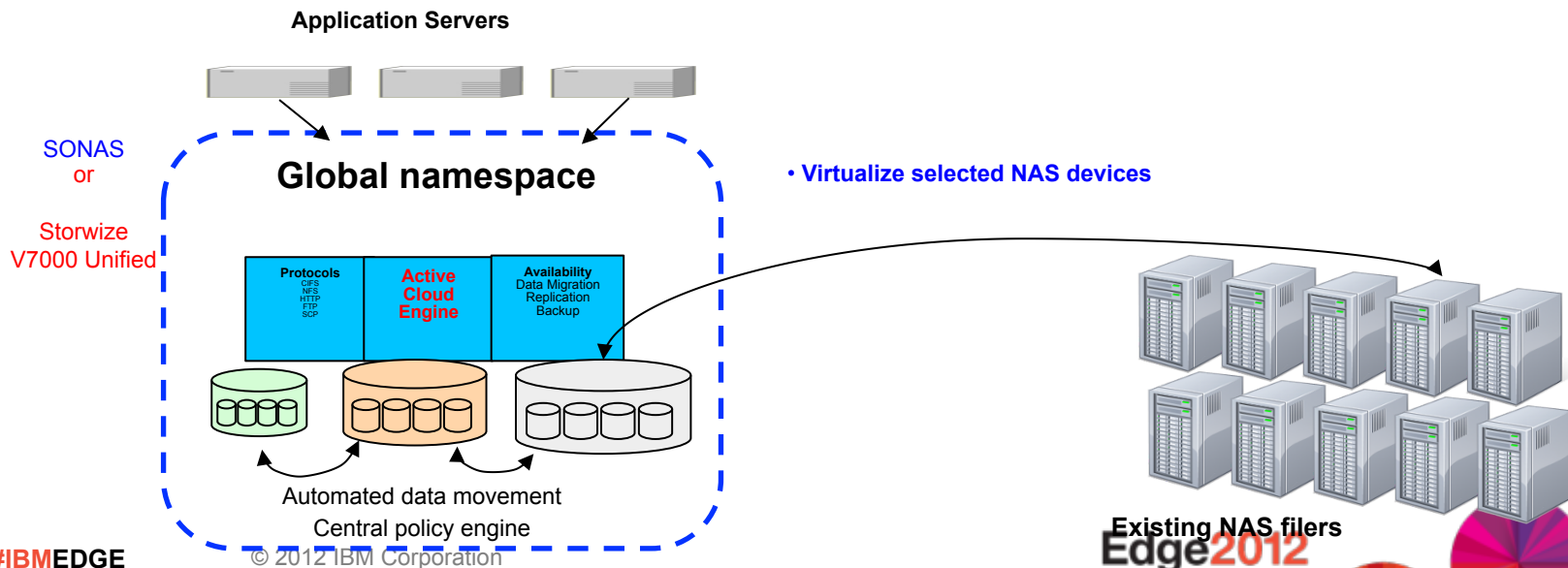
# IBM Active Cloud Engine

## Statements of Direction\*

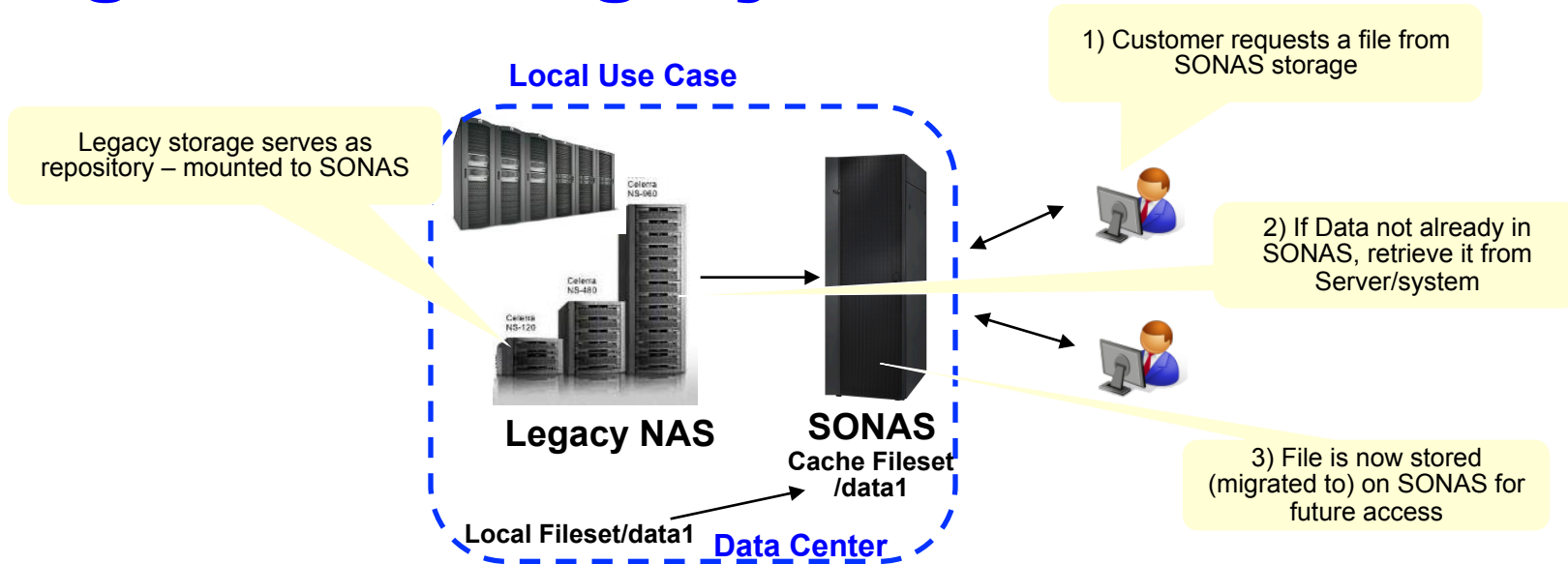


### IBM Active Cloud Engine™ will help provide:

- Virtualization capabilities while facilitating transparent file migrations from selected NAS systems to IBM SONAS and Storwize V7000 Unified
- Investment protection as clients can continue to leverage their existing NAS assets
- Improved productivity and true collaboration where users at geographically dispersed locations can both share and modify the same file.



# Statement of Direction\* - Migration of Legacy NAS data



## Definition

- Legacy NAS system is a “home site” to a SONAS “cache cluster”
- Migrate or access any data that Legacy NAS exports using NFSv3 protocol
- Will initially support read operations only

## Benefits

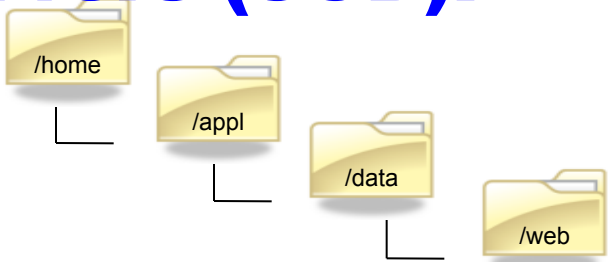
- Can provide transparent migration of data to SONAS or Storwize V7000 Unified while still allowing normal operations
- Reduces “down time” for users during migration period to just migrating ACLs, permissions, etc.



# Automated Data Migration from NFSv3 servers (SOD):



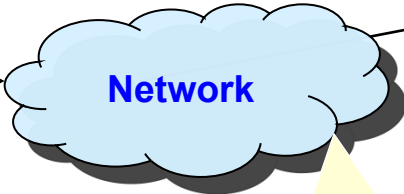
Insert SONAS ACE as front end - transparent background migration



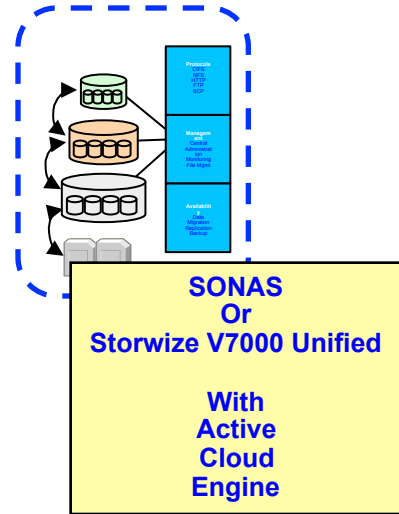
- /home/appl/data/web/important\_big\_spreadsheet.xls
- /home/appl/data/web/big\_architecture\_drawing.ppt
- /home/appl/data/web/unstructured\_big\_video.mpg



NFSv3 supported NAS server



Client opens a file  
SONAS cache cluster  
retrieves it from a NAS  
NFSv3 server



standard NFS

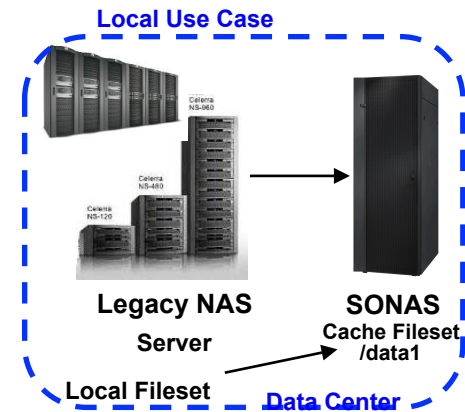
Pull in file on demand to SONAS infrastructure if needed. Can trigger a background prefetch for all



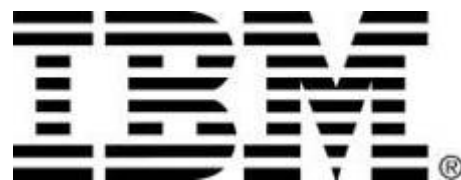
# Customer Example\*: Migration of Legacy NAS data using Active Cloud Engine

## Research Institute

- Existing NFS & CIFS environment, approximately 1PB and growing rapidly
- 10,000 students, employees/professors and associated research projects
- Mixed environment of NAS devices as well as approximately mixed protocols access via NFS only, CIFS only, and Mixed
- Original estimates: Over 6 months to migrate using current standard industry migration tools
- ACE Based Proposal for data migration
  - ACE can migrate data faster than industry standard tools
  - ACE can help migrate data from legacy NAS and servers via NFSv3 protocol
  - Goal is reduced overall time to migrate data and reduction of system down time
- **Future Lab Based Services Engagements required**



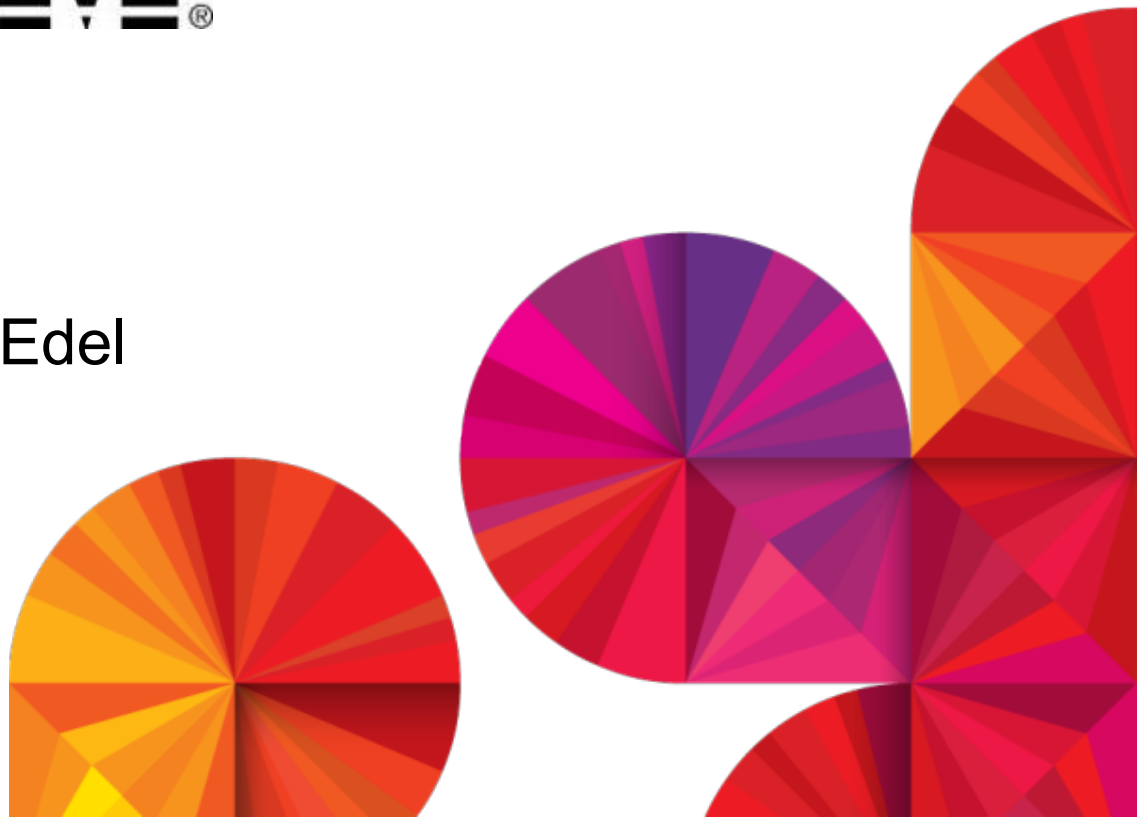
# Thank You!



Session: sVC32  
Presenter: Stephen Edel

**#IBMEDGE**

Intel, the Intel logo, Xeon and Xeon Inside are trademarks or registered trademarks of Intel Corporation in the U.S. and /or other countries.



# Trademarks and disclaimers

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries. IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency which is now part of the Office of Government Commerce. Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries. Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both. Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both. ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office. UNIX is a registered trademark of The Open Group in the United States and other countries. Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates. Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom. Linear Tape-Open, LTO, the LTO Logo, Ultrium, and the Ultrium logo are trademarks of HP, IBM Corp. and Quantum in the U.S. and other countries.

Other product and service names might be trademarks of IBM or other companies. Information is provided "AS IS" without warranty of any kind.

The customer examples described are presented as illustrations of how those customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics may vary by customer.

Information concerning non-IBM products was obtained from a supplier of these products, published announcement material, or other publicly available sources and does not constitute an endorsement of such products by IBM. Sources for non-IBM list prices and performance numbers are taken from publicly available information, including vendor announcements and vendor worldwide homepages. IBM has not tested these products and cannot confirm the accuracy of performance, capability, or any other claims related to non-IBM products. Questions on the capability of non-IBM products should be addressed to the supplier of those products.

All statements regarding IBM future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Some information addresses anticipated future capabilities. Such information is not intended as a definitive statement of a commitment to specific levels of performance, function or delivery schedules with respect to any future products. Such commitments are only made in IBM product announcements. The information is presented here to communicate IBM's current investment and development activities as a good faith effort to help with our customers' future planning.

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput or performance improvements equivalent to the ratios stated here.

Prices are suggested U.S. list prices and are subject to change without notice. Starting price may not include a hard drive, operating system or other features. Contact your IBM representative or Business Partner for the most current pricing in your geography.

Photographs shown may be engineering prototypes. Changes may be incorporated in production models.

© IBM Corporation 2012. All rights reserved.

References in this document to IBM products or services do not imply that IBM intends to make them available in every country.

Trademarks of International Business Machines Corporation in the United States, other countries, or both can be found on the World Wide Web at <http://www.ibm.com/legal/copytrade.shtml>.

