NUANCE

Spectrum Scale Transparent Cloud Tiering
Experiences at Nuance

Bob Oesterlin
Sr Principal Storage Engineer, Nuance

Nuance Company Overview

- Nuance Healthcare
  - Allows clinicians to capture and communicate care more naturally and efficiently while supporting better outcome
- Nuance Mobile
  - Creating a more human experience with cars, smart devices and IoT through voice, touch and natural language understanding innovations
- Nuance Enterprise
  - A more human, connected conversation customer service that integrates proactive engagement, voice biometrics, conversational systems, and multichannel virtual assistants
- Nuance Document Imaging
  - Document workflow and automation solutions that help customers to achieve measureable business and productivity benefits as they securely create, use and share documents
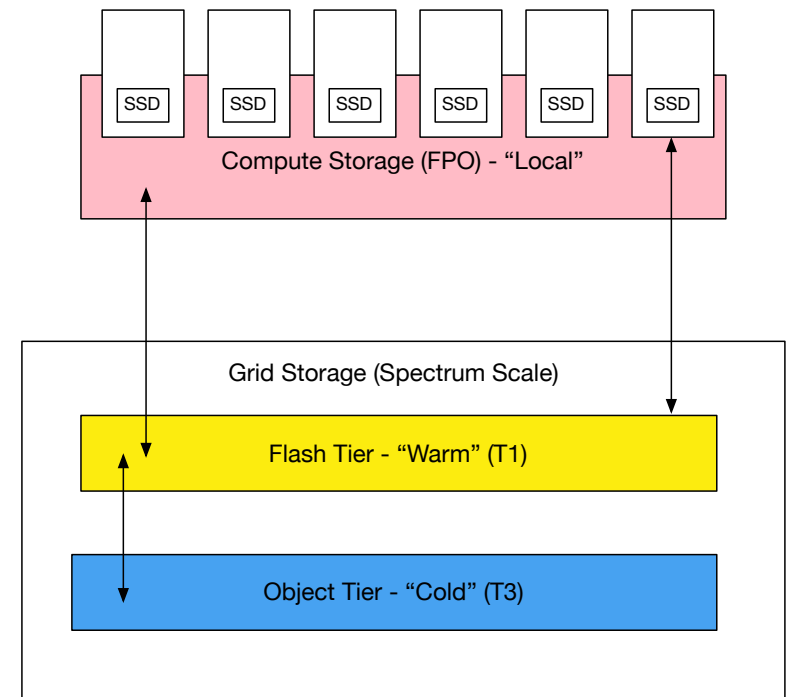
NUANCE

Nuance HPC Grids

- Supports the Worldwide Nuance R&D Community
- ~2000 TB per day of data processed
    - 90% Read, 10% Write
- ~2,000,000 jobs processed per day
- Over 4 PB of Spectrum Scale across multiple clusters
- On-premise Object Storage (Swift), 5 PB
- VMs for casual access/job submission

Big Data Challenges

- Continuous growth of data – billions of files
- Much of our data becomes cold quickly
    - As much as 80% of the data "cold" within a few weeks
- High performance storage is expensive
- Cloud/Object is the natural choice for colder data
    - But it's "yet another interface" for our developers
- Users have limited storage and data skills
- Predicting the future is hard

Transforming HPC Grid Storage

- Compute nodes have
  - Local SSD for high-speed R/O cache for use with SS
  - Local file systems between nodes
- Data flows freely between Compute, Local (FPO), and Warm Tier
- Data in T1 is on high-speed flash storage
- Data migrates based on policy from T1 to T3
  - Data prefetched from T3 to T1 "Tell us you need it"
  - Data migrated "on demand" from T3 to T1

SSD | SSD | SSD | SSD | SSD | SSD

Compute Storage (FPO) - "Local"

Grid Storage (Spectrum Scale)

Flash Tier - "Warm" (T1)

Object Tier - "Cold" (T3)

Spectrum Scale – Transparent Cloud Tiering

- Ideal solution for (Nuance) users to interface Cloud Storage
- Automatic or policy-based migration
- Data looks and feels the same
- Uses the existing Spectrum Scale CES Nodes (no longer a requirement in 4.2.2)
- Multiple choices for back-end cloud storage
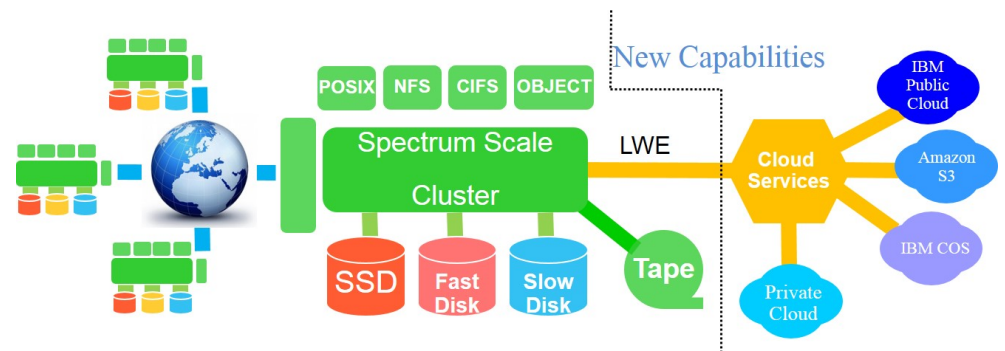
Transparent Cloud Tiering

- Nuance was an early beta tester
- Partnership with IBM to implement key features
- Early testing in 1H2016 proved technology
- Continued enhancements (performance, features) will be key to making deployment an overall success

- Requires Scale "Advanced" license
- Magic Config setting to enable use of TCT – ask IBM for it (I think this is crazy)

Transparent Cloud Tiering - Operational

- Object Tier appears as external storage pool
- Files migrated by mmapplypolicy, recalled by mmapplypolicy or on-demand

```
RULE 'OpenRead'
EVENT 'OPEN_READ'
ACTION(System('/opt/ibm/MCStore/bin/mcstore recall -c -i ' || varchar(INODE) || ' -g ' ||
varchar(GENERATION) || ' -s 0' || ' -f ' || varchar(FS_ID)) = 0)
WHERE(XATTR('dmapi.MCEA', 5, 1) == 'N')
RULE 'else' EVENT 'OPEN_READ' DIRECTORIES_PLUS

RULE 'OpenWrite'
EVENT 'OPEN_WRITE'
ACTION(System('/opt/ibm/MCStore/bin/mcstore recall -c -i ' || varchar(INODE) || ' -g ' ||
varchar(GENERATION) || ' -s 0' || ' -f ' || varchar(FS_ID)) = 0)
WHERE(XATTR('dmapi.MCEA', 5, 1) == 'N')
RULE 'else' EVENT 'OPEN_WRITE' DIRECTORIES_PLUS
```

Transparent Cloud Tiering - Operational

- Files exist in 3 states:
    - Resident: data blocks in the file tier only
    - Non-resident: data in cloud tier
    - Co-resident: data blocks in file/cloud tier

- Multiple version of files kept in cloud tier
    - Must be cleaned up by "reconcile" process
    - Process is slow
    - No (easy) way to determine overall cloud usage

NUANCE

Transparent Cloud Tiering - Operational

```
[root@cnw-r04r11u01 test]# echo This is a test > test.file
[root@cnw-r04r11u01 test]# dmls test.file
File name    : /gpfs/kam1/test/test.fileOn-line size : 15
Used blocks  : 0
Data Version : 0
Meta Version : 0
State        : ResidentBase
Name    : 0000000000000000.0000000000000000.0000000000000000.0000000000000000.00000000.0000000000000000
[root@cnw-r04r11u01 test]# dmmigrate test.file
[root@cnw-r04r11u01 test]# dmls test.file
File name    : /gpfs/kam1/test/test.file
On-line size : 15
Used blocks  : : Non-residentBase
Name    : 2336DADC14ACB58C.B87D327457A328FA.6E9A175836181E0A.0000000000000000.21648D16.00000000000CA7E3
[root@cnw-r04r11u01 test]# ls -l test.file
-rw-r--r-- 1 root root 15 Nov 13 09:37 test.file
[root@cnw-r04r11u01 test]# cat test.file
This is a test
[root@cnw-r04r11u01 test]# dmls test.file
File name    : /gpfs/kam1/test/test.file
On-line size : 15
Used blocks  : 32
Data Version : 1
Meta Version : 1
State        : Co-residentBase
Name    : 2336DADC14ACB58C.B87D327457A328FA.6E9A175836181E0A.0000000000000000.21648D16.00000000000CA7E3
```

Transparent Cloud Tiering - Experiences

- Performance is highly dependent of gateway node configurations
    - Dual 10 core system with 128GB memory is a good start
    - Tune GPFS config per documentation
- Make sure you Linux system (RH 7.2) is at the required version level – Hangs!
- Batched performance (via mmapplypolicy) is much higher than on-demand recall
- File system performance will impact migration throughput
- Move up to 4.2.2 if you can – features and performance (multiple filesystems, cloud tiers)
- Swift performance: Long file names, single container (2 objects per file)

- Sample performance numbers using 4-node gateway configuration
    - 20 core nodes, 160gb memory
    - 10gb interface
    - Flash storage for filesystem tier
    - Batched runs using mmapplypolicy

Transparent Cloud Tiering - Throughput

Test Type	Cleversafe (1gb Storage Nodes)		Cleversafe (10gb Storage Nodes)		Swift	
	Throughput	IOPs	Throughput	IOPs	Throughput	IOPs
Migrate 16,000 1meg files to cloud	614 MB/Sec	614	807 MB/Sec	808	289 MB/Sec	141
Recall 16,000 1meg files from cloud	829 MB/Sec	829	1205 MB/Sec	1230	1011 MB/Sec	1054
Migrate 16,000 2meg files to cloud	706 MB/Sec	355	1004 MB/Sec	501	294 MB/Sec	143
Recall 16,000 2meg files from cloud	921 MB/Sec	447	1302 MB/Sec	640	852 MB/Sec	416
Migrate 16,000 4meg files to cloud	738 MB/Sec	186	1025 MB/Sec	258	514 MB/Sec	143
Recall 16,000 4meg files from cloud	977 MB/Sec	246	1512 MB/Sec	380	1619 MB/Sec	416

NUANCE

Transparent Cloud Tiering - Wishlist

- Support for multiple cloud providers (part of 4.2.2)
- AFM Support
- Multiple container support per file system
- Reconcile Improvements
    - Based on number of versions, date/time (some improvements here in 4.2.2)
    - Policy-based (file type, fileset, etc)
    - Needs to be MUCH quicker
- Better tracking of cloud storage usage
- Performance improvements – scalability to billions of files

Thank you