

Data Reduction Attributes in nfsv4

draft-faibish-nfsv4-data-reduction-attributes-03

Sorin Faibish <faibish.sorin@dell.com>

Philip Shilane <philip.shilane@dell.com>

Motivation for Data Reduction (DR) attributes

- New storage arrays use expensive flash devices
 - Size of data sets is growing exponentially – exabytes
 - But size of backend flash storage grows linearly at higher price per GB
 - Storage servers compute power increases with use of much larger number of cores
 - Memory of the servers also increases using NVMe devices based memory (e.g., Intel Optane)
 - New faster NVMe over fabric interconnect is available (Intel)
 - New “NVM Express” and “NVM Express over Fabrics” protocols released

Motivation for Data Reduction (DR) attributes

- How we address this problem
 - New data reduction algorithms for deduplication and compression improve DR
 - Variable block deduplication improves 2-5x size of data on disk versus fixed block
 - New compression HW using MS zipline methods and/or Intel QAT chips are widely used in the industry
 - New DR require larger memories and large number of cores; new servers have both
 - What's missing: user information related to DR that arrays cannot know
 - The draft offers a way of transmitting DR attributes from NFS client to NFS server

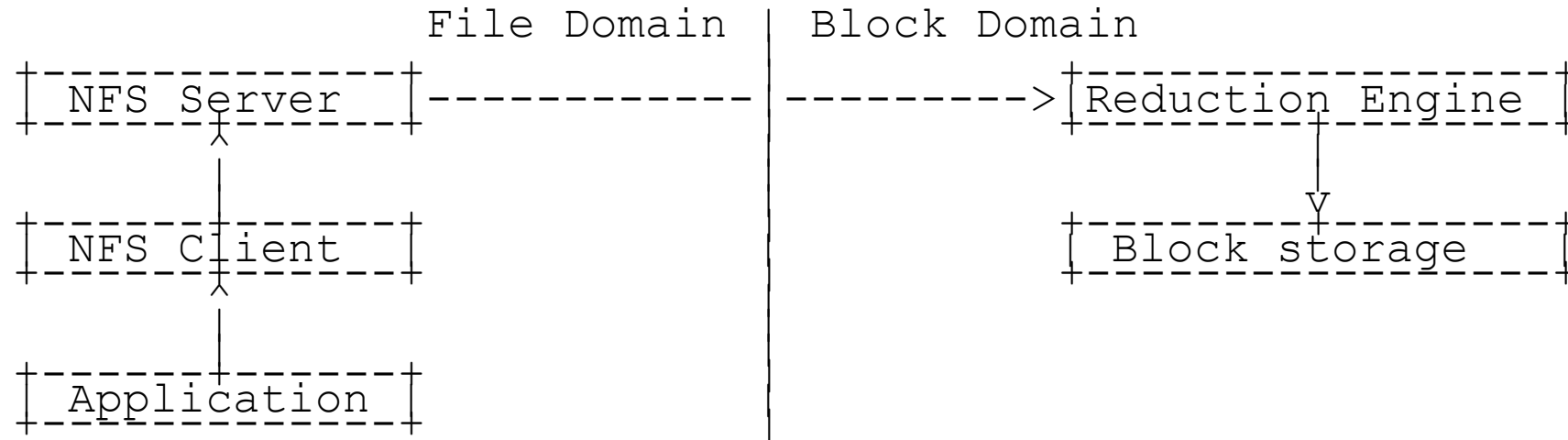
Solving the Problem (update)

- Currently NFSv4 has no means to communicate DR attributes to NFSv4 server
- NFSv4 server DR engine operates at FS block size typically 4k (Linux)
- There is analysts data regarding compression and dedupe of different types of files that can improve DR engines in the array efficiency
- There is no way to take advantage of this data as the application DR characteristics are not visible to the DR engine
- There is a new draft extending pNFS SCSI to NVMe-oF protocol that can be used (see new draft); we propose to introduce new DR Name Attributes
- We propose to expand the DR attributes to apply to pNFS SCSI

What is needed from NFSv4

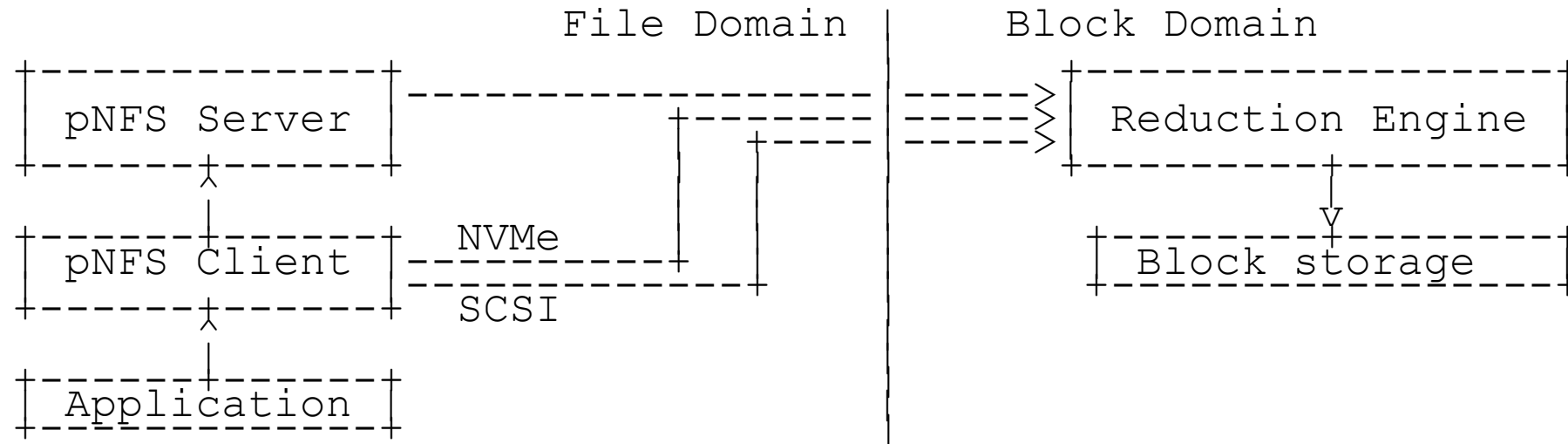
- We need a way to communicate DR characteristics from client to server
- We propose to add few new optional attributes to send compression and dedupe ratios from user/client to the NFSv4 server
- We do not propose any implementation solutions for DR; leave this to the NFSv4 vendor how to implement compression and deduplication
- We just allow better DR when the DR engine has additional information to target
- We assume that:
 - File extensions can indicate file type=> DR characteristics
 - File headers can indicate compression type (MPEG) => DR characteristics
 - Prevent applying DR on uncompressible data using file attributes => can use a flag

Use Case 1: NFSv4.2



- Assumptions for the use case:
 - NFS server can communicate RECOMMENDED attributes as metadata directly to the Reduction Engine
 - Backend identifies FS blocks of a “file” and can associate DR data to each block

Use Case 2: pNFS over NVMe/SCSI



- Assumptions for the use case:
 - DR named attributes are accessed by the OPENATTR operation using hidden directory of attributes associated with a file system object
 - NFS server extracts the data reduction RECOMMENDED attributes and pass their contents to the Block Storage

Typical Data Reduction Ratios

- Different applications have known DR/CR ratios:
(DR/CR = Data Reduction/Compression Reduction)

EDA	DR/CR=50%/50%
SWBUILD	DR/CR=0/80%
VDI	DR/CR=55%/70%
DB	DR/CR=0/50%
VDA	DR/CR=0/0
IT infrastucture	DR/CR=30%/50%
Oracle DW	DR/CR=15%/70%
Oracle OLTP	DR/CR=0%/65%
Exchange 2010	DR/CR=15%/35%
Geoseismic	DR/CR=3%/40%

New RECOMMENDED attributes

- Add new attributes associated to file system objects, e.g., files and directories.
- The RECOMMENDED attributes are stored with the file system objects
- SHOULD be preserved when files and directories are updated, moved or copied
- The attributes are hints from the client application regarding file compression and deduplication characteristics
- These attributes are intended for data needed by applications rather than by an NFS client implementation.
- NFS implementors are strongly encouraged to define the new data reduction attributes as RECOMMENDED attributes not mandatory
- The attributes are hidden metadata and SHOULD be retrieved by the NFS server and passed down to the data reduction engines of the Block Devices

Protocol Enhancements

- We propose enhancements to the NFSv4 protocol operations to allow DR RECOMMENDED attributes to be queried and modified by clients
- Add new attribute to bitmap4 data type to allow DR RECOMMENDED attributes support
- RECOMMENDED attributes may be examined and changed by normal GETATTR and SETATTR operations using the filehandles and stateid returned by OPEN command
- The Attributes can be modified by users and stored with the FS objects: files and directories.

Asks from NFSv4 WG

- Should DR attributes be added to the NFSv4 protocol as Named Attributes?
- Should this become a WG item?
- Should we first define the protocol changes to support DR before adoption?
- Is the WG interested in DR attributes as defined in this draft?
- We want to ask for WG review of the draft
- Next steps?

Future work

- We intend to write a new draft to allow hash keys exchange between NFSv4 server and client
- Main target to improve DR efficiency for cloud storage
- Is there any interest in the WG for this new protocol?