IN43B-1736

LVFS: A Big Data File Storage Bridge for the HPC Community

Navid Golpayegani[1], Dr. Milton Halem[2], Lucas Fonseca[3], Edward Masuoka[1]
navid.golpayegani@nasa.gov
[1] NASA Goddard Space Flight Center, Terrestrial Information Systems Lab
[2] University of Maryland Baltimore County, Department of Computer Science and Electrical Engineering
[3] Science Systems and Applications Inc.

Abstract

The objective of this work is to create a storage architecture that merges Big Data capabilities into the High Performance Computing's (HPC) architecture. This work concentrates on the storage architecture of the HPC community and how it can become scalable and flexible to manage current and future needs of the HPC community. The Lightweight Virtual File System (LVFS) utilizes the rich metadata information of a project's database to generate a dynamic directory structure. LVFS creates additional flexibility by creating separate and independent mechanisms for directory generating, content retrieval, and content access. The combination of these feature creates a powerful system to address the current and future big data needs for the HPC community.

For this work we consider two aspects of LVFS. First, LVFS has the capability to seamlessly merge existing and new hardware technologies, thereby giving the HPC system the ability to transition to new hardware technologies without disruption to the community's users. Second, LVFS addresses the need for new software technologies, such as differing file formats.

Introduction

LVFS is a plugin based storage system specializing in big data requirements. At its core, LVFS separates the most common services provided by a storage system, such as directory layout, file retrieval, and file access, into separate components. This separation gives LVFS the unique ability to add, remove, and update each component without disrupting the others. Each new component is developed as a plugin system for LVFS.

As an HPC system matures, switching to new hardware becomes more costly. For example, an HPC center supporting a project for many decades using standard block based storage disks will require a lot of code development to convert to utilizing a new block-based key-value storage system such as the one provided by Seagate's Kinetic Disks. With LVFS, a simple Kinetic Disk plugin for data retrieval can be developed and added. The existing directory plugins and data access plugins will utilize the new data retrieval plugin and retrieve files from the new Kinetic Disks in addition to the existing disks. The HPC center's users will be unaware of the new technology being used behind the scenes.

Another potential HPC issue is dealing with multiple file formats. Storing the same datasets in multiple file formats can be costly in both CPU and storage costs. With the LVFS plugin architecture, a combination of directory plugins and data access plugins allows for the support of arbitrary file formats of the same datasets without the need of additional disk space

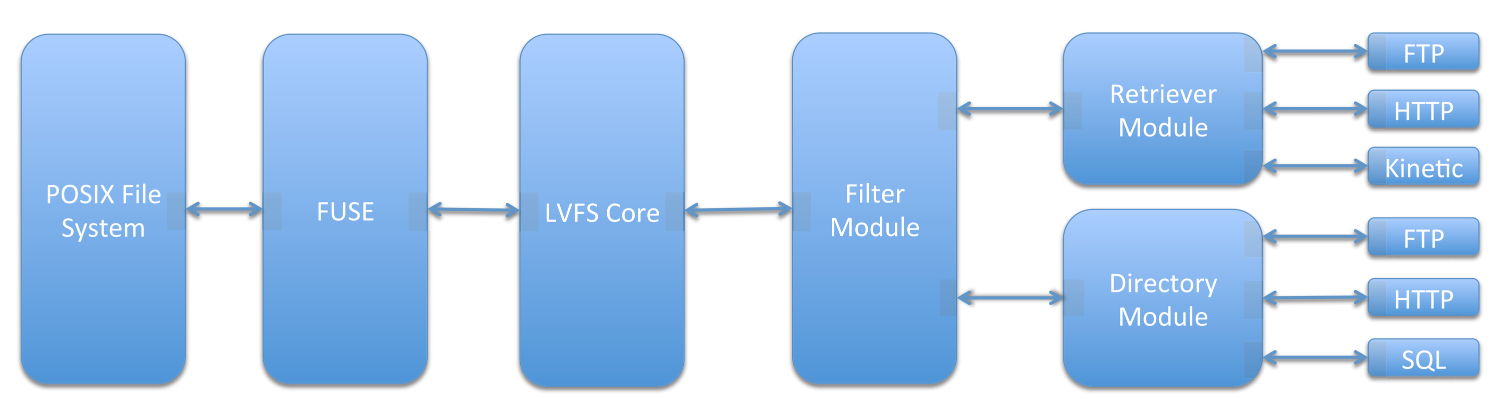
Technical details

Below is the technical details of the Bluewave cluster we will be using for this study. Bluewave consists of IBM's iDataPlex servers with both regular storage disks and Seagate's Kinetic Disk storage.

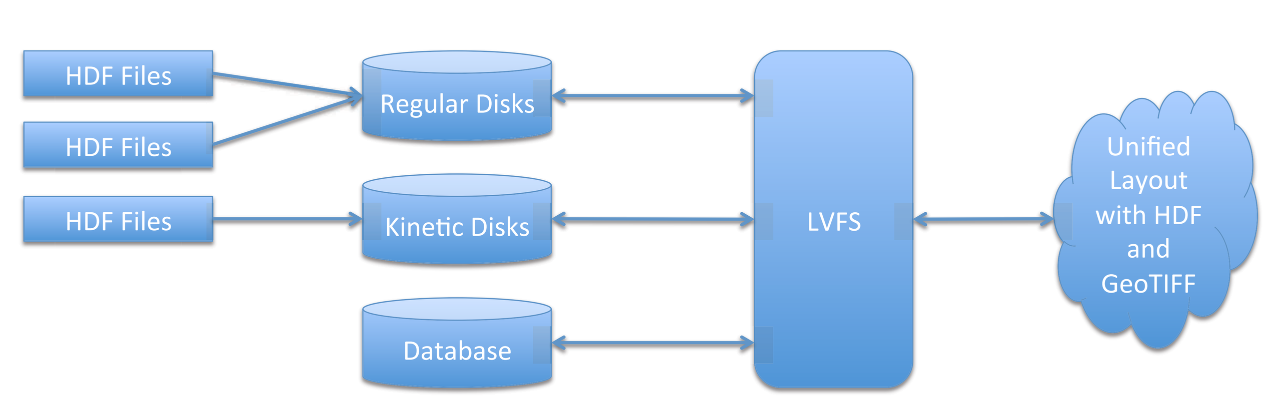
Bluewave
- 46.23 Tflop/s
- 4128 Total Cores
- 2 quad-core processors per node
- 24GB of RAM per node
- 2.8GHz Intel Xeon Nehalm (4 flop/s per clock)
- Infiniband DDR and 10Gig Ethernet

Implementation

Below is the stack overview of LVFS. The LVFS core provides the FUSE interface for POSIX compliance. POSIX calls are passed through the LVFS Filter plugins to the appropriate content or directory modules. The directory modules generate dynamic directory structures while the content modules retrieve data content. The Filter modules are responsible for modifying the results from the other modules. These modifications include creating new file listings or modifying the contents of a file.



The figure below shows an example how the LVFS plugin architecture for all three components generates a unified layout of HDF and GeoTIFF files. An LVFS plugin utilizes the Database server to generate a directory layout of all files available for the project hosted at the HPC center. A second plugin utilizes that information to determine if the actual HDF file is stored on a regular or kinetic disk and retrieves it from the appropriate location. Finally, a data access plugin converts the HDF files to GeoTIFF files. The end result is a unified view of both HDF and GeoTIFF files.



Testing and Evaluation

To test and evaluate these features we run LVFS both on the UMBC Bluewave cluster as a development instance and as a production system at NASA Goddard's Terrestrial Information Systems Lab for the LAADS data distribution system.

We used the UMBC cluster to evaluate the ability of LVFS to utilize new hardware technologies. To test the addition of new hardware we added Seagate's Kinetic storage simulator, an object-based key-value storage system. We were capable of storing files on both standard disks and add Kinetic disks without disruption to users. The directory layout continues to be the logical layout defined by the project while files are retrieved from regular or kinetic disks as needed.

The LAADS distribution system was used for testing the generation of GeoTIFF files of the Level 3 gridded HDF files. Existing HDF files are fed into LVFS which, through a filter module, would generate corresponding GeoTIFF files. Certain HDF files, like the MOD08_D3, contain hundreds HDF SDSs and, as a result, take approximately thirty seconds to generate. To alleviate performance issues, LVFS is capable of letting users read a GeoTIFF file while it is being generated ensuring that only blocks that are finalized are read. This allows for processing to start nearly immediately after the GeoTIFF conversion starts rather than requiring a thirty second delay.

Conclusion & Future Work

This work presents a unique approach to creating a big data storage environment that is capable of handling future needs of the HPC community. By separating directory generation from data retrieval and data access LVFS has the ability to create a unified layout utilizing both old and new hardware thereby creating no disruption the user's view with no costly requirement of moving files from old disk to new. Additionally, the ability to modify data during data access, LVFS is capable of supporting newer file formats without requiring additional storage needs by the HPC center. Users are capable of reading files in nearly any format supported via LVFS plugins. With this design, an HPC center is capable of supporting current and future projects with identical features all while allowing individual projects to utilize standard POSIX storage functions.

LVFS is currently running as a production system for the LAADS data distribution. The on-demand file format conversion mechanism is planned to be deployed for production in early 2016 and expanded to more file formats, such as NetCDF.

Navid Golpayegani
NASA Goddard Space Flight Cente
Code 619
8800 Greenbelt Road
Greenbelt, MD 20771

navid.golpayegani@nasa.gov
(301) 614-6781