QuickSilver: A Distributed Policy Engine for Lustre

ORNL Software and Data Expo (OSDX'22)

Chris Brumgard, Anjus George, Rick Mohr, Ketan Maheshwari, James Simmons, Sarp Oral

May 10-11, 2022

`gitlab.ccs.ornl.gov/techint/policy-engine`

Overview

QuickSilver is a light-weight, flexible and distributed policy engine to manage data on a tiered Lustre filesystem. QuickSilver is composed of single-purpose agents that handle tasks such as gathering file metadata, enforcing policy decisions, and executing policy actions like purging or data migration.

These agents are designed to communicate using the NATS message queue software, allowing the number of individual agents to be scaled up as needed. QuickSilver is designed to function while maintaining minimal state information.

Motivation

- Frontier Orion Lustre filesystem is composed of two tiers:
  - An SSD-based Performance Tier
  - An HDD-based Capacity Tier
- Need for a software solution to manage data across these tiers
- Solution must be scalable and reliable
- Scalability ensures it could operate from multiple nodes in a distributed fashion
- Reliability ensures continued Lustre operations in the face of increased IO volume and spikes

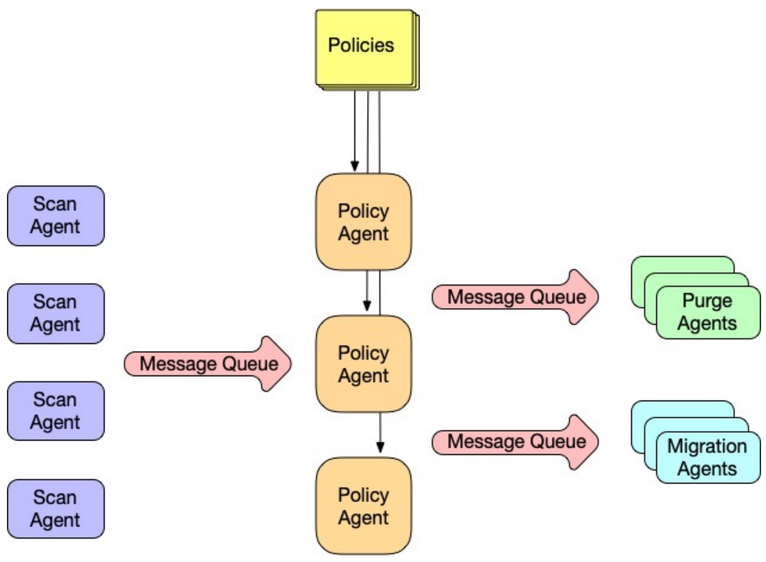
Design

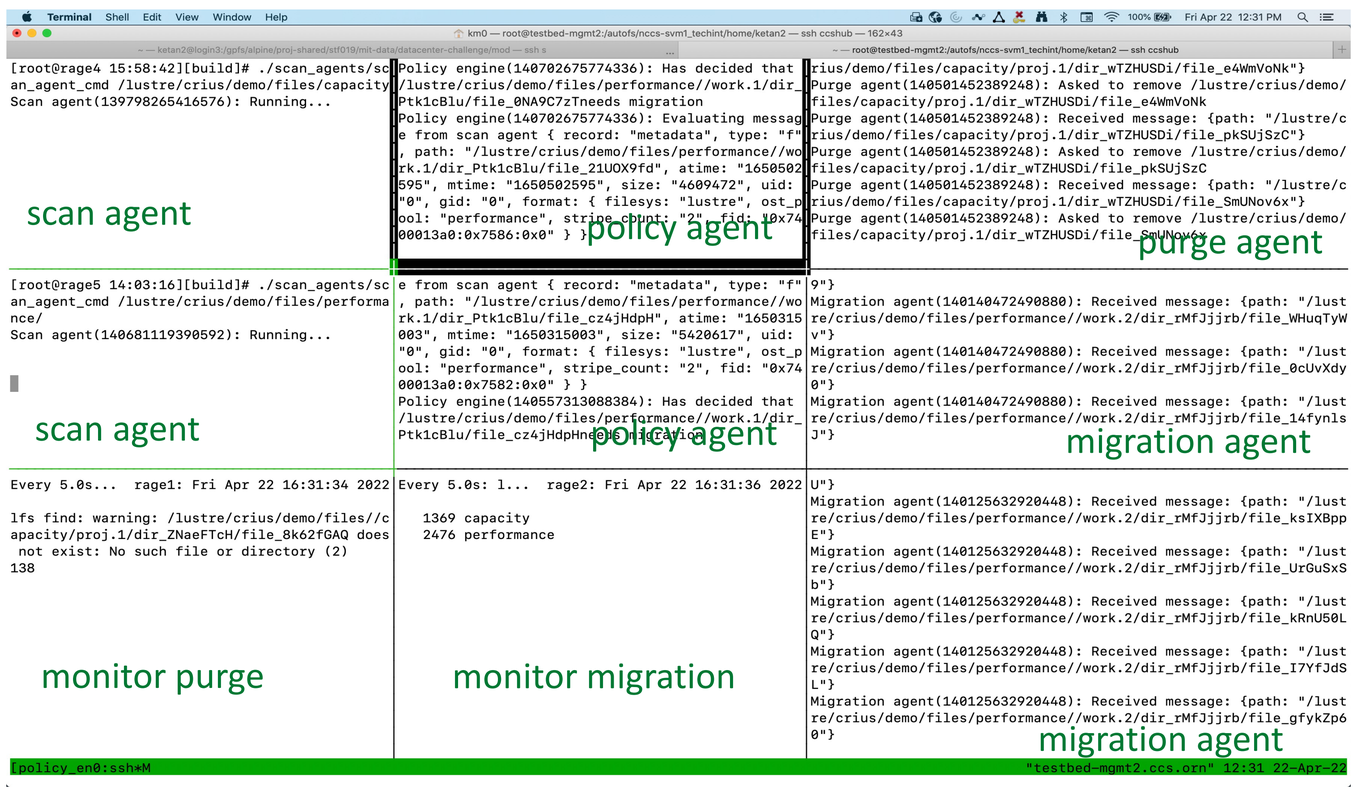
- Comprised of fully independent and distributed components
- Scan Agents collects the filesystem state using Lustre's lfs find
- Policy Agents processes the collected filesystem metadata and interprets predefined policy definitions
- Purge Agents performs purge actions on files eligible for purging
- A last-minute check done per file before purge action
- Migration Agents performs migration operation on files eligible for migration

Tools and Techniques

- A decoupled agents-based publish-subscribe style implementation using the C++17
- Uses the NATS message queuing library [https://nats.io]
- NATS serves as a coordination and message queuing system
- Multiple NATS servers enable resiliency, high availability and load-balance in case of high traffic volume

Architecture and Implementation





Summary

- A distributed and scalable policy engine to manage data on a tiered Lustre filesystem. Designed as independent agents working in coordination.
- Implemented using C++ platform using NATS message queue system.
- An on-going development with a working prototype. Performance of the current implementation measured over a 1 million files mock dataset.
- The system works in a fully distributed mode over distinct independent nodes

Acknowledgments

This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725. We would like to thank our collaborators Dustin B. Leverman, Jesse A. Hanley, Bran Radovanovic, Philip Curtis.

OAK RIDGE National Laboratory

U.S. DEPARTMENT OF ENERGY