OAK RIDGE National Laboratory

PoliMOR: In Action

Chris Brumgard

brumgardcd@ornl.gov

LUG 2024

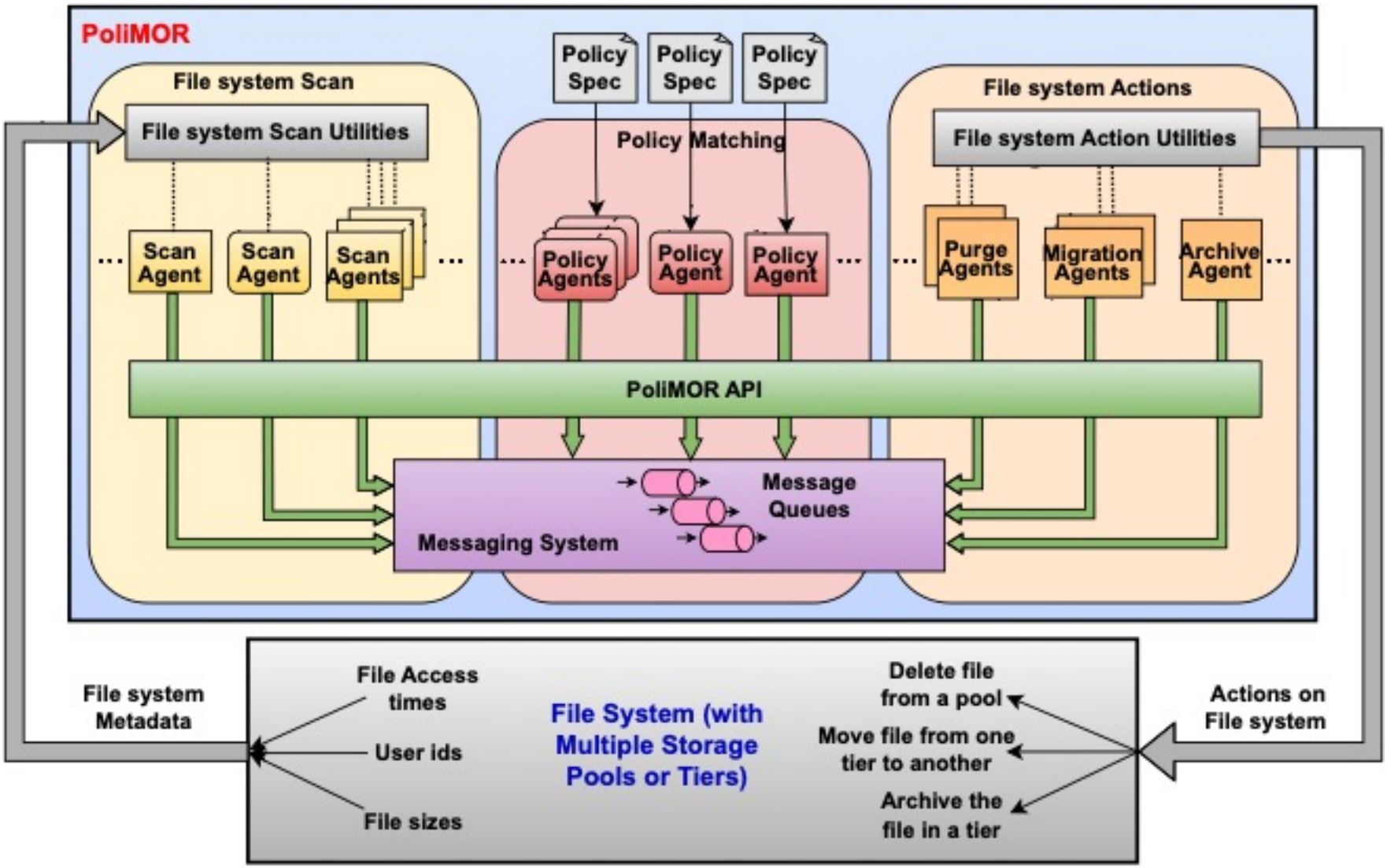
ORNL is managed by UT-Battelle LLC for the US Department of Energy

U.S. DEPARTMENT OF ENERGY

PoliMOR

- A Distributed, Extensible, and Automated Policy Engine.
- Purposes:
  - Migration, Purging, Data Collection, and Telemetry for Lustre.
- What do we mean by Distributed?
  - Agents connected via a distributed messaging queue system.
  - Fault Tolerance & Scalable.
- What do we mean by Extensible?
  - Agent-Oriented so new agents can be added for functionality.
- What do we mean by Automated?
  - Define a set of invariants within a policy to be maintained.
  - No intervention by the users or admins.

PoliMOR diagram



PoliMOR

- At previous LUGs, we've discussed:
  - The need for PoliMOR
    - Migration between tiers
    - Provide admins an easy and reliable way to implement policies for different users/groups
  - The design of PoliMOR
    - Agent-Based, Scalable, Distributed, Fault-Tolerant, Extensible
  - Preliminary results
- Discussing actual deployment data from Orion in live production.

Production Setup

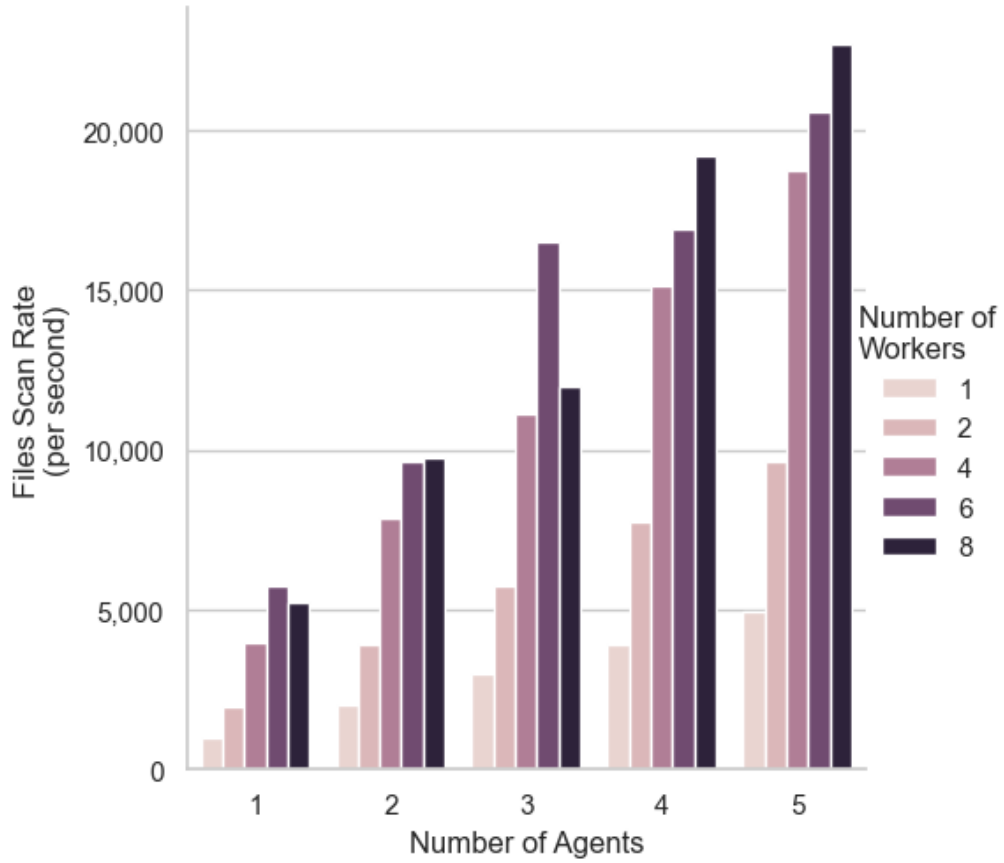
- Orion
  - 40 MDS nodes (9.7 PB NVMe metadata tier)
  - 450 OSS nodes (10 PB NVMe performance tier and 679 PB capacity tier)
  - Cray Custom Lustre version 2.15
  - Slingshot

- Orion Util nodes
  - 9 nodes AMD EPYC 7402P 24-Core Processor, 128 GB of RAM
  - Slingshot

Agent Scaling Tests

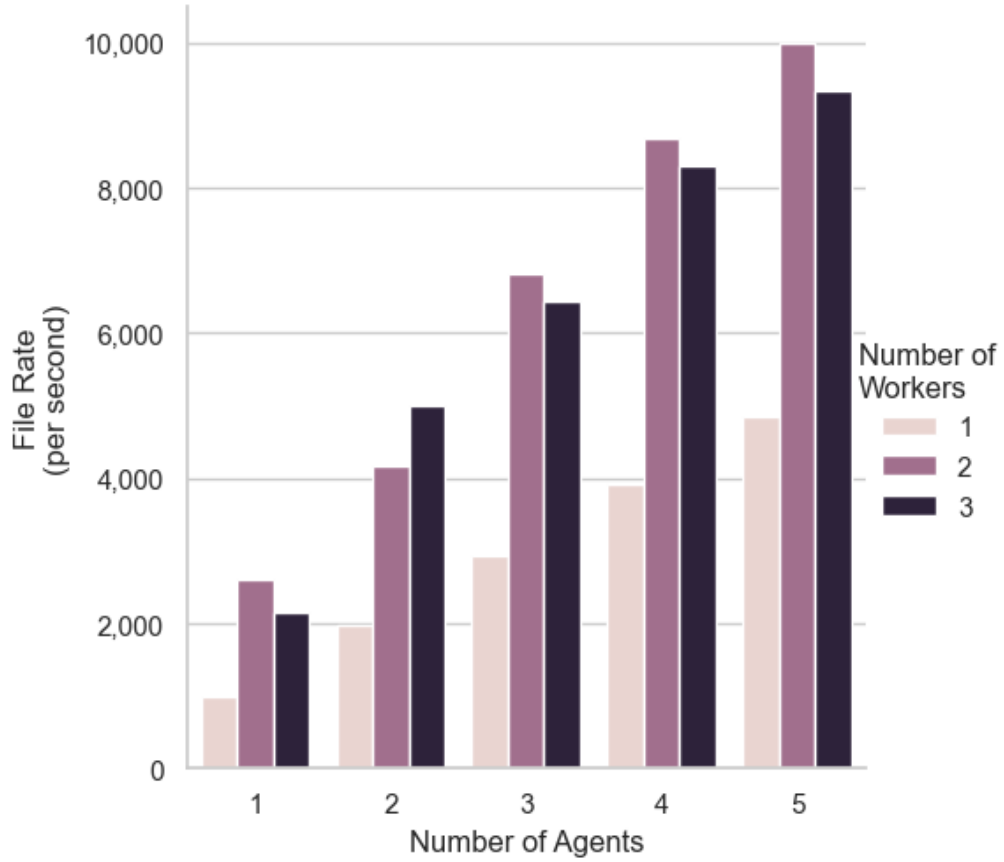
- Test the scaling of each agent.
- Synthetic Project Sandbox on Orion.
  - ~11.7 million files spread across a directory tree of 27,700 directories.
  - All files are purgeable.
  - Files are uniformly randomly sized from 1 B to 2 MiB.
    - The first 256 KiB are on the MDT (DOM)
    - 256 KiB to 1 MiB are performance-tier OSTs
    - 1M+ is on capacity-tier OSTs

Agent Scaling: Scan Agent



- Near-perfect linear scaling as the number of agents increases.
- Performance tends to improve per agent as workers increase.
  - Sometimes breaks down after 6 workers.

Agent Scaling: Policy Agent

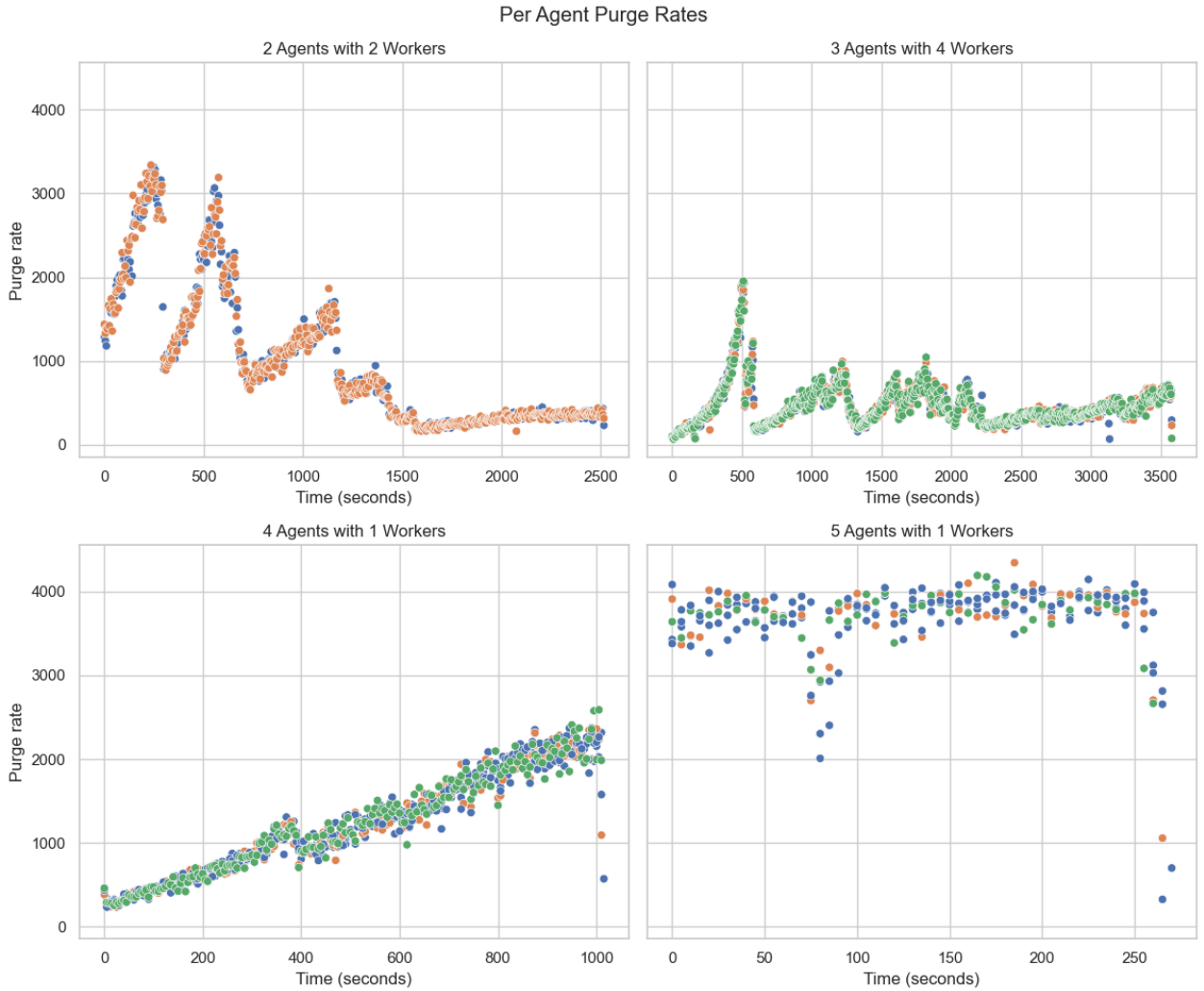


- Perfect linear scaling as the number of agents increases.
- Optimum number of workers is 2.
  - Each worker waits on a file to arrive on the input queue, processes it, and then sends the result on the output queues.

Agent Scaling: Purge Agent

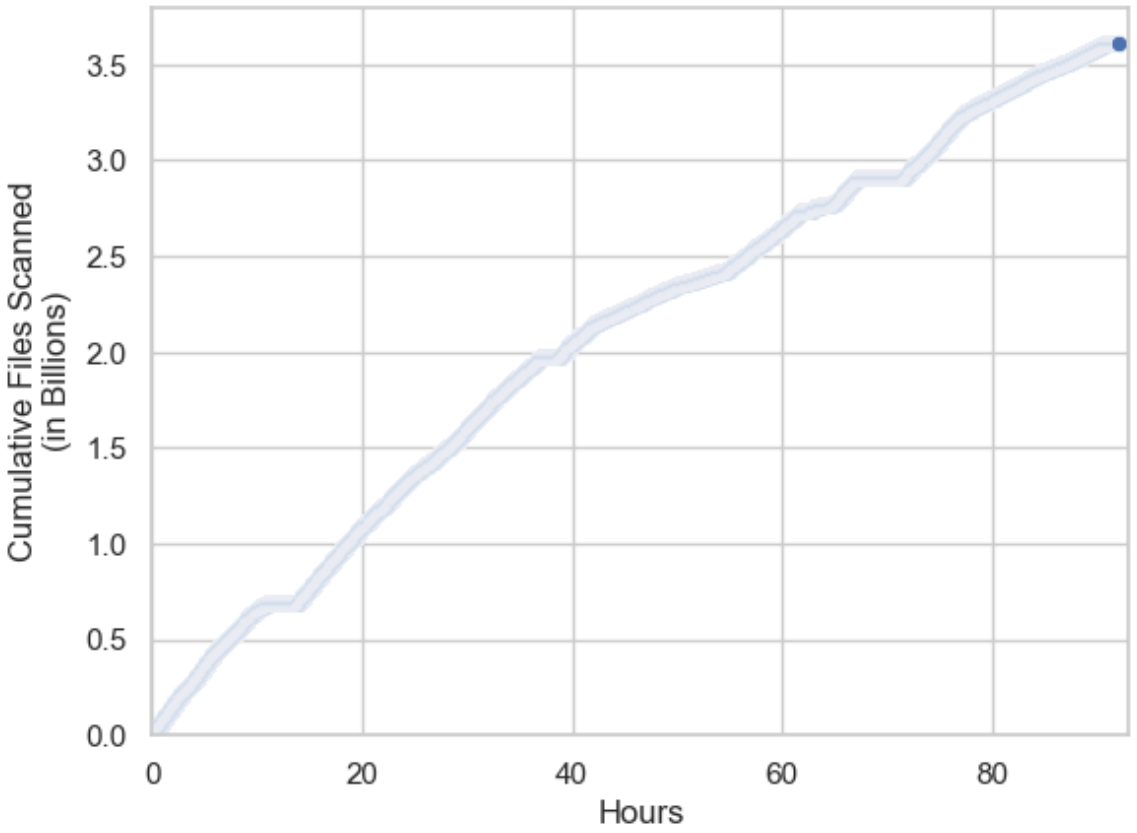
- Better than linear scaling for agents.
- Number of workers is not linear.
- Includes full synchronous stats to verify timestamps.

Agent Scaling: Purge Agent



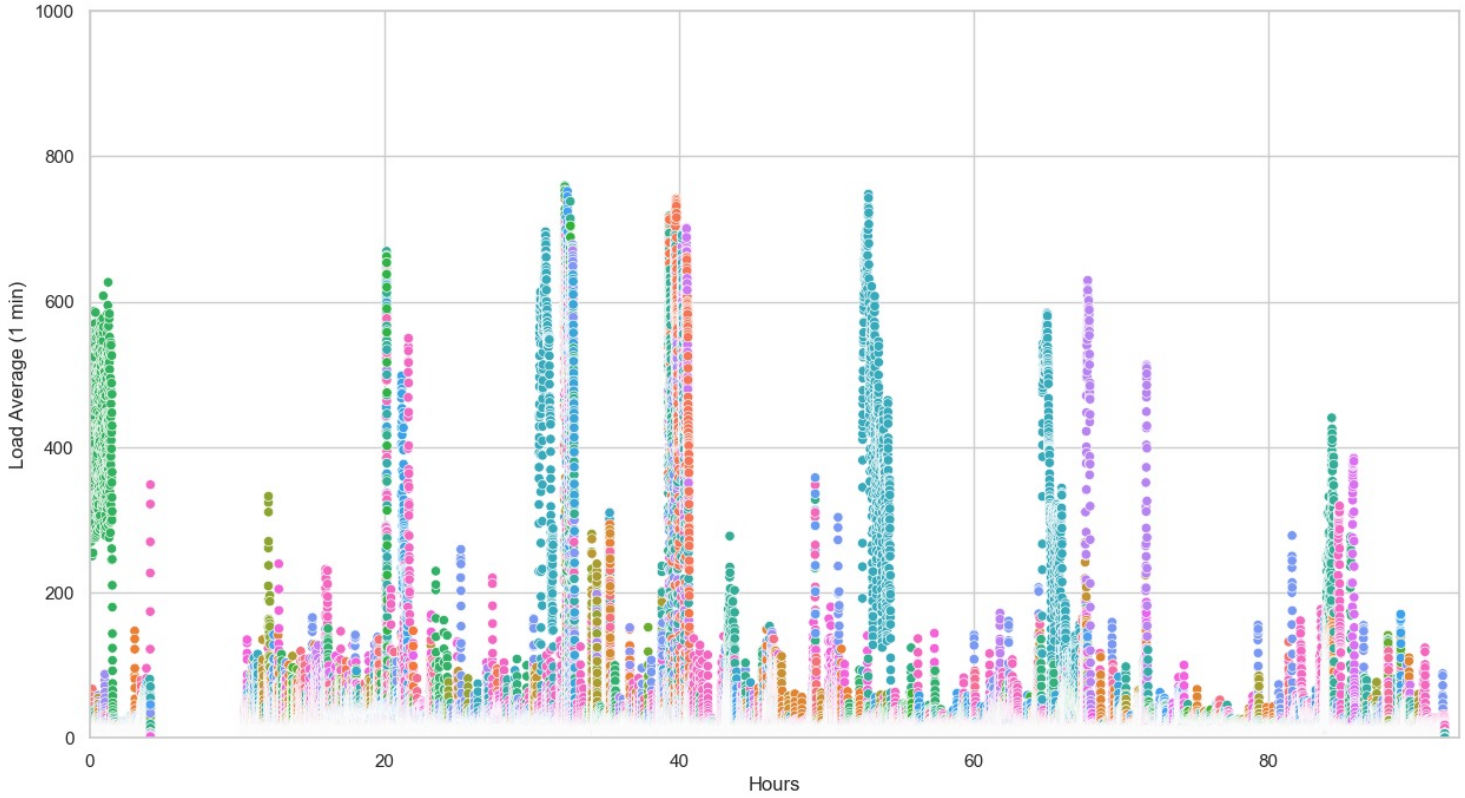
- For 3 or fewer agents, saw-tooth behavior for each agent (different color).
  - Tends settle with a low tail.
  - Every agent (separate node) has the same pattern.
- For 4 or more agents, this behavior starts disappearing.
  - See either a ramp-up or just steady behavior.
- Doesn't appear to be LRU lock (count or age) or a memory issue.
- More investigation is needed.

Entire Orion Filesystem Walk



- 5 scan agents and 10 policy agents.
- Scanned and processed 3.62 billion files in 92 hours!!!
- Several issues were encountered:
  - 1st and 3rd flat spots
    - NATS servers locked up
  - 2nd flat spot
    - MGS and Util nodes issue
  - Slowdowns
    - An agent or two would stop getting NATS messages.
- Despite issues, PoliMOR always made forward progress and completed!!!

Entire Orion Filesystem Walk



MDS Server Load Average

- During run
  - Mean 5.2 @ Std Dev. 25.5
  - Median 0.27
- Normal
  - Mean 5.2 @ Std Dev. 44.5
  - Median 0.23

What's Next?

- Full production onto Orion.
- Supervisor tasks to handle the agents and NATS servers.
- NATS usage improvements.
- Scan Agents
  - Improve the work-sharing between the scan agents.
  - Reducing the scan work.
- Several potential optimizations in the purge and migration agents.
  - Directory caching and directory-aware functions.
  - Cluster work by directories per agent.

Future work

- More complex policies.
  - Decomposing complex actions into simpler rules.
- Non-Lustre agents.
  - HPSS
  - Edge and Streaming
- Looking for collaborators.

Acknowledgments

- Thanks to my fellow developers Anjus George, Ketan Maheshwari, Rick Mohr, James Simmons.
- 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.