# DAOS System Administration

## System Monitoring

System monitoring and telemetry data will be provided as part of the control plane and will be documented in a future revision.

## System Operations

### Full Shutdown and Restart

Details on how to support proper DAOS server shutdown will be provided in future revision.

### Fault Domain Maintenance & Reintegration

Details on how to drain an individual storage node or fault domain (e.g. rack) in preparation for maintenance activity and how to reintegrate it will be provided in future revision.

### DAOS System Extension

Ability to add new DAOS server instances to a pre-existing DAOS system will be documented in future revision.

## Fault Management

DAOS relies on massively distributed single-ported storage. Each target is thus effectively a single point of failure. DAOS achieves availability and durability of both data and metadata by providing redundancy across targets in different fault domains.

### Fault Detection & Isolation

DAOS servers are monitored within a DAOS system through a gossip-based protocol called SWIM[[1]](#footnote-2) that provides accurate, efficient and scalable server fault detection.

Storage attached to each DAOS target is monitored through periodic local health assessment. Whenever a local storage I/O error is returned to the DAOS server, an internal health check procedure will be called automatically. This procedure makes an overall health assessment by analyzing the IO error code and device SMART/Health data. If the result is negative, the target will be marked as faulty, and further I/Os to this target will be rejected and re-routed.

Once detected, the faulty target or servers (effectively a set of targets) must be excluded from each pool membership. This process is triggered either manually by the administrator or automatically (see next section for more information). Upon exclusion from the pool map, each target starts the collective rebuild process automatically to restore data redundancy. The rebuild process is designed to operate online while servers continue to process incoming I/O operations from applications.

Tools to monitor and manage rebuild are still under development.

### Rebuild Throttling

The rebuild process may consume a lot of resources on each server and can be throttled to reduce impact on application performance. This current logic relies on CPU cycles on the storage nodes. By default, the rebuild process is configured to consume up to 30% of the CPU cycles, leaving the other 70% for regular I/O operations.

During rebuild process, the user can set the throttle to guarantee the rebuild will not use more resource than the user setting. The user can only set the CPU cycle for now. For example, if the user set the throttle to 50, then the rebuild will at most use 50% of CPU cycle to do rebuild job. The default rebuild throttle for CPU cycle is 30. This parameter can be changed via the daos\_mgmt\_set\_params() API call and will be eventually available through the management tools.

## Software Upgrade

Interoperability in DAOS is handled via protocol and schema versioning for persistent data structures. Further instructions on how to manage DAOS software upgrades will be provided in future revision.

### Protocol Interoperability

Limited protocol interoperability is provided by the DAOS storage stack. Version compatibility checks will be performed to verify that:

* All targets in the same pool run the same protocol version.
* Client libraries linked with the application may be up to one protocol version older than the targets.

If a protocol version mismatch is detected among storage targets in the same pool, the entire DAOS system will fail to start up and will report failure to the control API. Similarly, the connection from clients running a protocol version incompatible with the targets will return an error.

### Persistent Schema Compatibility and Update

The schema of persistent data structures may evolve from time to time to fix bugs, add new optimizations or support new features. To that end, the persistent data structures support schema versioning.

Upgrading the schema version will not be performed automatically and must be initiated by the administrator. A dedicated upgrade tool will be provided to upgrade the schema version to the latest one. All targets in the same pool must have the same schema version. Version checks are performed at system initialization time to enforce this constraint.

To limit the validation matrix, each new DAOS release will be published with a list of supported schema versions. To run with the new DAOS release, administrators will then need to upgrade the DAOS system to one of the supported schema versions. New pool shards will always be formatted with the latest version. This versioning schema only applies to a data structure stored in persistent memory and not to block storage that only stores user data with no metadata.

## Storage Scrubbing

Support for end-to-end data integrity is planned for DAOS v1.2 and background checksum scrubbing for v2.2. Once available, those functionality will be documented here.

1. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1028914 [↑](#footnote-ref-2)