



Starfish v4

Release Notes



Starfish Storage Corporation ("COMPANY") CONFIDENTIAL
Unpublished Copyright (c) 2011-2018 Starfish Storage Corporation, All Rights Reserved.

NOTICE: All information contained herein is, and remains the property of COMPANY. The intellectual and technical concepts contained herein are proprietary to COMPANY and may be covered by U.S. and Foreign Patents, patents in process, and are protected by trade secret or copyright law. Dissemination of this information or reproduction of this material is strictly forbidden unless prior written permission is obtained from COMPANY. Access to the source code contained herein is hereby forbidden to anyone except current COMPANY employees, managers or contractors who have executed Confidentiality and Non-disclosure agreements explicitly covering such access. ANY REPRODUCTION, COPYING, MODIFICATION, DISTRIBUTION, PUBLIC PERFORMANCE, OR PUBLIC DISPLAY OF OR THROUGH USE OF THIS SOURCE CODE WITHOUT THE EXPRESS WRITTEN CONSENT OF COMPANY IS STRICTLY PROHIBITED, AND IN VIOLATION OF APPLICABLE LAWS AND INTERNATIONAL TREATIES. THE RECEIPT OR POSSESSION OF THIS SOURCE CODE AND/OR RELATED INFORMATION DOES NOT CONVEY OR IMPLY ANY RIGHTS TO REPRODUCE, DISCLOSE OR DISTRIBUTE ITS CONTENTS, OR TO MANUFACTURE, USE, OR SELL ANYTHING THAT IT MAY DESCRIBE, IN WHOLE OR IN PART. FOR U.S. GOVERNMENT CUSTOMERS REGARDING THIS DOCUMENTATION/SOFTWARE These notices shall be marked on any reproduction of this data, in whole or in part. NOTICE: Notwithstanding any other lease or license that may pertain to, or accompany the delivery of, this computer software, the rights of the Government regarding its use, reproduction and disclosure are as set forth in Section 52.227-19 of the FARS Computer Software-Restricted Rights clause. RESTRICTED RIGHTS NOTICE: Use, duplication, or disclosure by the Government is subject to the restrictions as set forth in subparagraph (c)(1)(ii) of the Rights in Technical Data and Computer Software clause at DFARS 52.227-7013. Unpublished - All rights reserved under the Copyright Laws of the United States.

Please report corrections to Support@StarfishStorage.com



Table of Contents

Version History

[Version 4.0.4785 \(9/04/2018\)](#)
[Version 4.0.4686 \(7/24/2018\)](#)
[Version 4.0.4622 \(7/3/2018\)](#)
[Version 4.0.4608 \(6/27/2018\)](#)
[Version 4.0.4564 \(6/14/2018\)](#)
[Version 4.0.4501 \(5/24/2018\)](#)
[Version 4.0.4387 \(4/17/2018\)](#)
[Version 4.0.4351 \(4/9/2018\)](#)
[Version 4.0.4277 \(3/16/2018\)](#)
[Version 4.0.4218 \(2/27/2018\)](#)
[Version 4.0.4181 \(2/12/2018\)](#)
[Version 4.0.4139 \(1/31/2018\)](#)
[Version 4.0.4084 \(1/17/2018\)](#)
[Version 4.0.4011 \(12/20/2017\)](#)
[Version 4.0.97 build 3978 \(12/12/2017\)](#)
[Version 4.0.95 build 3919 \(11/28/2017\)](#)
[Version 4.0.94 build 3847 \(11/7/2017\)](#)
[Version 4.0.93 build 3832 \(10/31/2017\)](#)
[Version 4.0.92 build 3788 \(10/25/2017\)](#)
[Version 4.0.91 build 3768 \(10/17/2017\)](#)
[Version 4.0.90 build 3741 \(10/9/2017\)](#)
[Version 0.6.89: No release](#)
[Version 0.6.88 build 3705 \(9/26/2017\)](#)
[Version 0.6.87 build 3686 \(9/19/2017\)](#)
[Version 0.6.86 build 3664 \(9/11/2017\)](#)
[Version 0.6.85 build 3632 \(9/5/2017\)](#)
[Version 0.6.84 build 3598 \(8/28/2017\)](#)
[Version 0.6.83: No release](#)
[Version 0.6.82: No release](#)
[Version 0.6.81 build 3542 \(8/14/2017\)](#)
[Version 0.6.81 build 3516 \(7/31/2017\)](#)
[Version 0.6.80 build 3473 \(7/19/2017\)](#)
[Version 0.6.79 build 3444 \(7/12/2017\)](#)



[Version 0.6.78 build 3424 \(7/5/2017\)](#)

[Version 0.6.77 no release](#)

[Version 0.6.76 build 3383\(6/23/2017\)](#)

[Version 0.6.75 build 3354 \(6/14/2017\)](#)

[Version 0.6.74 build 3325 \(6/7/2017\)](#)

[Version 0.6.73 build 3279 \(5/30/2017\)](#)

[Version 0.6.72 build 3239 \(5/23/2017\)](#)

[Version 0.6.71 build 3235 \(5/19/2017\)](#)

[Version 0.6.70 build 3188 \(5/5/2017\)](#)

[Upgrade Process](#)

[Other Notes](#)



Version History

Version 4.0.4785 (9/04/2018)

Fixes

- Tier Stats query (in redash Tier Dashboard) had invalid %used calculation.
- Improve error messages in Starfish cli when querying using invalid regex
- Improvements in Starfish client autocomplete response time
- sf cron list failing after new volume added, prior to that volume completing first scan. sf cron list will now only display volumes that completed first scan.
- mtime scan failed due to individual directories that took longer than 1 hour to get file list from
- Improvements in mtime performance
- Automatic heartbeat added to uid/gid resolution phase of scan, to avoid situation where long resolution timed out and failed scan.
- Volume add using invalid volume name format responds with clearer error
- Lustre event monitor utf-8 compatibility fix for pgloader.
- During Starfish installation, Sentry did not install even though it was selected to do so.
- GUI:
 - Filters now have effect in action mode (only items matching the search & filters will have the action taken on them)
 - Recover mode properly colors a directory if it has contents in it that is recoverable
 - Release notes now downloadable via the life-preserver button
 - Going to jobs tab and back to browser tab now works as expected
 - Search box was sometimes contracted when it should have been expanded
 - Clean up formatting and add "last accessed date" to file details
 - Show all versions now working for deleted files
 - In recover mode, directories now properly colored with their recoverability (dir color based on the directories file contents and whether they are archived)
 - Prevent opening removed volumes after a refresh
 - Disable archiving (plus open in new tab, copy dir path) when more than one item selected (you can archive all files matching a filter or search but only one arbitrary item (file or dir) can be selected and archived)
 - Correctly show the source and destination paths for archive jobs



- In search pane, make location selector clearer (current dir, or volumes / all volumes)
- Performance improvements and smaller bug fixes

New Features

- Minor log enhancement when adding new Starfish volume.
- `sf scan show` and `sf scan list --detail` enhanced to show Starfish volume and path
- Temporary query result files automatic clean up added
- Added ability to query/report on SFID. This uniquely identifies a specific version of a file in the Starfish database.
- Added "--quiet" option to "`sf scan start --wait`" command.
- Updated Postgres database settings to better leverage parallelism and fast disks.
- `sf collect-logs` now supported on agents
- Automatic obfuscation added to `sf collect-logs` for archive targets
- "sf job issues" featured added. Requires sentry to be installed.
- Archive manifests directory name changed from "manifests" to ".sfarchivemanifests"
- Archive reporting updates:
 - Only show "Phase" while archive or recovery is running
 - Display all job-level errors in archive/restore job output. Previously only one error was displayed
- GUI:
 - New right-click menu command to clear all selected items (for action mode, to take an action on all items in the browser pane)
 - New right-click menu for a bundle of files in tree view and flat view browser panes (beta)
 - Added license status indicator in upper right of GUI
 - New version available notification in the upper right corner
 - New delete action (with explanation and dry run option) in all places actions can be requested (beta)
 - Added regular expression searching to various search conditions
 - In show-CLI-query log (under "eyeball" details menu button), allow entering of a CLI-formatted query and show the result in the GUI (beta)
 - In the details pane of a file or directory, job results are shown, and clicking on the job result header switches to the job tab, showing the job details and the list of files affected by the job
 - Tooltip delay can be set to other than the default 2 second delay (in `/opt/starfish/etc/99-local.ini: tooltip_popup_delay_seconds=<# of seconds>`)



- Volumes in the volume pane can be sorted name, size, and free space, also filtered by glob expression
- "Latest copied version" now an advanced search condition (show the latest copy in any archive target)
- New tags tab, showing all tags and per-tag details (beta)

Upgrade Notes

- Schema updated from version 62 to 64. This is a minor schema upgrade.

Version 4.0.4686 (7/24/2018)

Fixes

- Removed "query" option from API, as duplicate of "queries"
- Archive started from GUI was producing incorrect estimates, causing job to show "PARTIAL_SUCCESS"
- CLI responsiveness improvements
- Event monitor pause enhancements. This relates to Lustre monitor agents where Starfish agent daemon is killed before stopping properly. Monitor status now pauses properly, allowing monitor to be continued instead of having to stop and restart.
- Increase Starfish agent daemon stop timeout from 15 to 30 seconds.
- sf check-config: added explicit error when WAL size is exceeded. This occurs when archiving falls behind in a configuration that uses database backup and/or replication.
- GUI:
 - Action mode properly works on all items in browser pane if nothing is selected (Note there is still a bug that filters are ignored in action mode browser pane, should be fixed in next release)
 - A variety of small bug fixes
 - Internally, many actions are now asynchronous, allowing commands to have greater effect and canceling of actions to cancel the resulting action within the core (i.e. a canceled search causes the core to stop working on the search)
 - Archive and recover now properly save and recreate a directory tree when used recursively
 - Allow only 1 item to be selected for archiving at a time (a file or a dir tree)

New Features



- Added number of threads option to sf scan start. This allows one to select number of concurrent threads to run when manually starting a scan.
- sf volume list: now can use multiple volumes as argument, instead of a single volume.
- sf query: now can sort files by depth. Note that --sort-by requires --type d or --type f option.
- GUI:
 - Beta of PAM integration for Starfish administrative access
 - Hashing action promoted to GA
 - “Show CLI Query Log” toggle under “more options” icon promoted to GA - shows equivalent CLI command for many actions within the GUI (searching, filtering, browsing)
 - Treemap make dirs non-transparent unless being moused-over (opposite of how it worked previously)
 - Add archive-target has dst_path field to allow adding a preface to files/objects copied there
 - Tagging a volume (via the edit volume pane) - tag applies to the entire volume
 - Recovered archive items can optionally be recovered to location other than where they were archived from
 - Jobs tab now shows which files/dirs the job affected (i.e. what files were archived in an archive job)
 - Recover icon now more differentiated

Upgrade Notes

- None

Version 4.0.4622 (7/3/2018)

Fixes

- Query was not properly recognizing that --not --type d was the same as --type f.
- Monitor updated to wait for crawler processes to exit before exiting itself.
- Archive job error messaging expanded for job failures.
- GUI:
 - Fix tree view to filter appropriately.
 - Add back in “scan” in right-click menu.

New Features

- Job error messages updated
- GUI:



- For access time and modification time filters, add a config option to ignore directory dates or file dates. That allows a filter to be set, and even if a directory's date doesn't match the filter, the directory to be show to allow browsing into its contents (which to match the filter), for example.
- Allow recovery (beta) right-click menu item if search condition "has-copy" is included (i.e. the search is showing copies available on archive targets, so allow them to be recovered).
- User and Configuration Guide updated with archive and recover details (beta).

Upgrade Notes

- Schema updated from version 60 to 62. This is a minor schema upgrade.

Version 4.0.4608 (6/27/2018)

Fixes

- Recover timeout
- Traceback when running "sf check-config" and agent version does not match server version.
- Removed Sentry from installer.
- GUI:
 - Fixed extraneous pop up scan errors

New Features

- Install and Configuration guides updated. These are accessible via Starfish GUI.
- Added target path to archive. This allows archiving to a subdirectory of an archive-target.
- GUI:
 - In compare-2-dates mode, new (hidden by default) columns for size change and count change, allowing sorting by, for example, biggest size change.
 - Added size and count information in treemap view current directory
 - New toggle in directory size filter to have it apply just to a directory or recursively (the directory tree)
 - Add selectable dedupe mode to archive action (for archive beta testers)
 - Add right-click context menu to bundles of files in list and tree view (beta, disabled by default)
 - Disable PAM authorization by default (in beta, not ready for GA)



Version 4.0.4564 (6/14/2018)

Fixes

- Pending queries are now cancelled when Starfish services are cycled.
- sf tag show updated to correctly fail when only argument is a colon
- volume size numbers changed to more clearly delineate value: 'M' changed to 'MB', 'Ti' changed to 'TiB'
- Workaround for pgbackrest not installing on CentOS 6
- GUI:
 - Improved recover mode (for archive beta users)
 - Improved performance of sunburst & treemap views
 - Sunburst now shows user-IDs if there is no user name translation available
 - Search warning that a search might take a while now appears more appropriately
 - Various small bug fixes and performance improvements

New Features

- tags now support "/", "?", and "." characters
- Starfish now includes /etc/pam.d/starfish, which is configurable via auth.pam_service_file property in Starfish configuration. Access to Starfish requires credentials plus membership of group configured in auth.required_group_name. If the user is a member of auth.admin_group_name, they also get superuser access.
- Install and Configuration guides updated. These are accessible via Starfish GUI.
- GUI:
 - Email to support or about a POC extension now buttons under the help ("life preserver") icon
 - Improved labels on volumes
 - Made treemap view non-transparent by default, with mouse-over making a box transparent (to see the contents of the represented directory)
 - Improved data-as-of calendar date selection
 - Made it more obvious what mode the GUI is in (for example, history or recovery mode)
 - Added an indicator that a scan is running and a refresh option when the scan is done
 - Volume filtering is now regex format (i.e. "t*" for every volume name that starts with a "t")
 - Indicator that a new version of Starfish is available
 - Right-click action to show all version information about a file
 - Toggle to hide or show hidden (".*") files
 - Toggle for search field to be case sensitive or case insensitive



- Improved compare 2 dates mode

Version 4.0.4501 (5/24/2018)

Fixes

- Workers_per_agent not being properly used when starting job.
- Increased timeout for postgres metrics. These are queried by sf check-config and the daily email home report.
- `http://sfserver/redash` now redirects to `https://sfserver/redash`
- `sf query --perm /o+w` no longer includes symlinks in output
- GUI version 1.0.2:
 - Performance improvements and minor bug fixes
 - Scan and job status and results now match CLI information
 - Individual files now show count=1 instead of count=1, sizes, aggregates now accurate

New Features

- New volumes automatically update capacity once a day.
- Existing volumes can be set to update capacity once a day.
 - Run '`sf volume set-total-capacity volname auto`' to modify
- `sf collect-log` now includes nginx logs
- Agent install now defaults to https
- UTF-8 locale now required for Starfish server and agents.
- Case insensitive query for groupname and username added.
- Postgres replication status added to `sf check-config/daily` report
- Updates to `sf scan pending --detail` for lustre event monitor
- Event monitor pause stops crawlers immediately
- GUI version 1.0.2:
 - Volume pane now condensable and resizable
 - Sunburst and treemap functionality improvements
 - “Goto Directory” added to “more” button dropdown menu
 - Advanced search now part of the search pane, with many new advanced search options, standard search now allows selection of where the search should take place (current dir, selected volume(s), all volumes)
 - Jobs, scans tabs now scrollable with resizable panes



- Hashing as action and right-click action (beta - disabled by default)
- Archiving (beta - disabled by default)
- Read-only mode (set "read_only=true" in [GUI] section of /opt/starfish/etc/99-local.ini)
- Tagging allowed in read-only mode (set "allow_tagging=true" in [GUI] section of /opt/starfish/etc/99-local.ini)
- New "atime" column in flat and tree browsing modes
- Hash and other job details and results added to details panel in tree and flat browsing modes
- More tooltips (popup after mouse hover for 2 seconds)
- In details, where there is a hash value for a file, click the value to search for other files with a matching value (duplicates)
- Mode bar in recover, history, compare 2 dates modes to make more obvious the current mode

Upgrade Notes

- New dependencies: iotop, lshw
- Starfish will not start if non-UTF-8 locale is in use

Version 4.0.4387 (4/17/2018)

Fixes

- Improved Event Monitor performance statistics in 'sf scan pending' command
- sf volume list was failing to show when cron service was down
- redash information dashboard column named changed from 'timezone' to 'start time'
- duplicate_check error when using --check-unique-file-size fixed
- GUI version 1.0.1:
 - Change ctime to mean change time rather than create time
 - manage tags -> add tag in sunburst view hung with spinning disk

New Features

- Added listing of huge directories to show_failed_entries.sh script
- Beta development of Lustre Event Monitor
- GUI:
 - Show the official version number in the CLI header
 - Improved the tool tip for "data as of"
 - Added "open Starfish in new Tab" option when showing the "cancel" button



- In compare 2 dates mode, option to only show what has changed (and not show unchanged (white) items)
- For 504 timeout errors, display whatever debugging information we can
- Add mtime scan to right-click menu (to rescan a specific directory branch)
- Show current directory info in all browser views
- Added a "reset to latest data" button to the bottom of the "data as of pane"
- Remove "directory size" and "age" from advanced search (do those via filters instead)
- Added "depth" condition to advanced search
- Added "permission" condition to advanced search

Upgrade Notes

- No schema updates

Version 4.0.4351 (4/9/2018)

Fixes

- sync-and-verify timeout increased
- API link in Starfish GUI was redirecting to incorrect location
- Slowness in 'sf job list' resolved
- Event monitor not killing child processes properly when monitor is stopped
- Fixed initial scan failure occasionally seen
- Set Analyze to not run if autovacuum was running. For busier sites that usually have autovacuum running, this will allow autovacuum to complete return of dirty/expired pages to database for reuse
- Help updates for recover/restore
- GUI:
 - ctime is "change time" not "create time" - fixed in details pane
 - Found some redundant API calls, removed to improve performance
 - Comparing 2 dates in history now correctly color codes removed, new, and changed
 - Also colors now reset when leaving compare 2 dates mode
 - In recover mode, properly color coding what is available to recover
 - Updated "i" info pane with the details
 - Fixed colors going to black in sunburst and treemap views
 - Cancel button now works properly during long request
 - Changed colors of elements to make it clearer when something is available vs. grayed out
 - When filtering, aggregates can be inaccurate on dirs so grayed out, but still accurate on files so not grayed out



- "Data as of" information now accurately shows when the scan ran that updated the currently viewed information
- If UID to name isn't available, now shows UID instead of blank
- Scan metrics (number of files and dirs scanned) now accurate
- Add archive-target type S3 field URL is no longer required (now optional)
- Removed trailing "\$" for cost in details pane
- If a file was hashed (via sf template command) sometimes the hash wouldn't show in the details pane
- If doing a subsequent search, the summary of the previous search no longer appears while the search is running



New Features

- Improvements in scan statistics, particularly when using Starfish event monitor
- Improvements in scan failure cause description
- Minimum depth query added to sf query
- Cause of failure added to archive and restore jobs
- GUI 1.0.1 released
 - Can now type in a date (in the format of yyyy-mm-dd) in the calendar picker pane
 - read-only mode now available to disallow all changes (including tagging) via the "read_only=True" flag set in 99-local.ini
 - In read-only mode, enable tagging but no other actions (such as archiving) via the "allow_tagging=True" flag in 99-local.ini
 - "type" conditional added to advanced search
 - A search that might take a while (because it is also filtered, or it is unindexed) now has a pop up window allowing cancel, run in another tab, or continue
 - Instead of "read more" buttons in jobs and scans tabs, now a draggable pane divider
 - Several color / font / look improvements
 - Under "More options" button, new option to "sort dirs first" to put directories on top and files at the bottom, no matter how columns are sorted.

Upgrade Notes

- Schema updated from version 58 to 59
- The "Upgrade Process" at the end of this document has been updated and clarified.

Version 4.0.4277 (3/16/2018)

Fixes

- Failure to move chunk caused scanner to fail. Fix is to delete the chunk once processed if it can't be moved for short-term retention.
- Event monitor pause command now waits until pause is completed before returning to prompt.
- Failure of archive and restore jobs include additional details on failure cause.
- Fixed sort order in redash for %delta and cost per gb in tag and user change reports
- PGLoader failed due to directory deeper than approximately 1000. After the fix all paths no longer than kernel's PATH_MAX characters are supported. Most systems have PATH_MAX set to 4096.
- Event monitor was not clearing failure messages in sf scan status properly. This is now fixed.
- End date on redash User Size Diff report changed to last scan, from current date.



- Various event monitor updates for Lustre Event Monitor beta

New Features

- Starfish GUI general availability
 - Starfish GUI is available at <https://<sfserver>>
 - Login name is "starfish"
 - Password is the secret_key from /opt/starfish/etc/01-general.ini on Starfish server
 - Will be automatically installed and updated going forward.
 - See below for feature overview
- Option to automatically update volume capacity periodically added
 - Run 'sf volume set-total-capacity <volumename> auto' to enable.
- Code optimizations for super-dense directories. We also added an option to skip these directories and log an error; the default is to log an error on directories with more than a million files, and skip directories with 2 million or more objects. Both of these can be modified.
 - We will continue to explore ways to improve performance for single directories with several million files, but in general it is better to work with users to change behavior. Most NAS/filesystem vendors recommend against having super-dense directories.
- During new install, only use UTF-8 encoding in Starfish database.
- Crawler optimization for large directories.

Upgrade Notes

- Schema version upgrade from version 56 to 58

GUI Features

We are pleased and excited to announce the first GA release of the Starfish GUI.

The GUI is an HTML 5 interface to the Starfish core and database. It has been tested against the following browsers:

- Chrome 64.0.3282.186
- Firefox 57.0
- Safari 11.0

High-level features

- Four main tabs: Dashboard (a subset of Re:dash reports), Browser, Jobs, and Scans
- Browsing of the metadata (including POSIX and Starfish tags) of all configured volumes in the Starfish database (including details)
 - Aggregate size, count and cost information (aggregated per directory branch)
 - Tree, flat, sunburst, and treeview browsing modes
- History/version viewing including a point in time and differences between two points in time
- Globbed searching of the metadata
- Advanced search including:



- across all volumes
 - via conditions including name, extension, age, tag, dir size, UID, GID
- Filtering based on access date/time, modification date/time, file size and directory size and per-column
- Edit volume information, including cost
- Create/edit/delete/search tags
- Actions include Export to CSV
- Help includes API, User/Configuration, and Installation documentation, tooltips, and an "info" button to explain mode operation

Version 4.0.4218 (2/27/2018)

Fixes

- Archive service returns more informative errors when same command run twice
- Archive service support for --no-prescan option fixed
- Width of sf archive list --detail set to 140 characters

New Features

- Help file updates
- From address for daily email can now be customized.
 - This is set in /etc/sysconfig/starfish using the "DIAGNOSTIC_REPORTS_FROM_EMAIL" variable.
- Lustre event monitor beta released

Upgrade Notes

- No Schema updates

Version 4.0.4181 (2/12/2018)

Fixes

- Exception handling of UTF non-Unicode/invalid code points (code points that are not encodable as Unicode in UTF-16, and thus not storable as UTF-8)
- archive-target addition of volume failed when dst_path was not specified.

New Features

- sf query --versions and --pit features work together



- API updates
- Help file updates

Upgrade Notes

- No Schema updates

Version 4.0.4139 (1/31/2018)

Fixes

- Treat symlink target as broken if null.
- Various updates to Starfish CLI help/man pages
- Crawler optimized to use significantly less RAM when dealing with a dense directory, such as 10s of millions of files in a single directory.

New Features

- archive-target verifies availability of archive target when definition created or updated.
- sf recover command added as alias for sf restore command
- sf collect-logs now captures archive and restore logs.
- sf archive command now supports volume, in addition to object storage
- "show" output of sf archive, sf restore, and sf archive-target commands updated to be similar to sf scan show.
- Documentation now included with Starfish, and available through Starfish GUI.
 - Also available in /opt/starfish/gui
- Documentation update to 1-30-2017 version.

Upgrade Notes

- Schema version upgrade from version 54 to 56

Version 4.0.4084 (1/17/2018)

Fixes

- Scan failures on directory stubs fixed.
- Optimization in postgres query plans for new volume scans. Postgres was selecting a non-optimized query.
- Removed User size over time pulldown reports from User Global analytics dashboard. They were causing redash logs to grow huge, and were duplicated by User Detail analytics dashboard.



- Fix column name in Re:dash volume report.
- Duplicate check timeouts:
 - Timing out during tagging phase. Reduced size of batch and from 1000 to 200 and added --tag-batch-size option to duplicate_check command.
 - Timing out during tag purging phase. Optimized query responsible.

New Features

- Added sf restore list and show
- Added verification of metadata phase of restore from archive
- Added "has a copy" field. This will help find files that do/do not have an archive copy.
- Added "on-disk" query parameter.
- Infrastructure updates for Lustre Event Monitor
- Updated various indexes (job_results_history, tag_results tables)
- Optimizations to reduce memory footprint for scanning.

Upgrade Notes

- Schema version upgrade from version 49 to 53

Version 4.0.4011 (12/20/2017)

Note: Starting with this version, numbering has changed from "4.0.xx.buildnumber" to "4.0.buildnumber".

Fixes

- Initial scan failures fixed by increasing timeout.
- Speed improvements in event monitor resolver

New Features

- Daily email notifications now send errors only to end-user. Errors and warnings continue to be sent to sf-status@starfishstorage.com
 - In /etc/sysconfig/starfish, email addresses on the "DIAGNOSTIC_REPORTS_EXTRA_EMAILS" line are end-user emails, which now include errors only.
 - In /etc/cron.d/starfish, email addresses on the "report_stats.sh" line receive errors and warnings.
- sf archive and sf restore now include "start" verb to start an archive or restore, similar to sf job verbiage.



- sf archive-target command parameters updated from json to item=value pairs
- sf query adds --whole-history and --job-id query options.
- help expanded for various commands.

Upgrade Notes

- Schema version upgrade from version 47 to 48

Version 4.0.97 build 3978 (12/12/2017)

Fixes

- re:dash verbage clarification in volume global analytics dashboard: Updated "directories without files" to "Directories with subdirectories only (no files)"
- redis was failing to start in IPv4 only environment. No longer bind redis to IPv6.
- Job results without volume failed during previous schema 45 upgrade
- Duplicate checker timeouts; switched to bulk-tagging method and increased timeout
- Hardlink failures in event monitor
- redash tag reports ignore archive tags

New Features

- Archive targets now are a combination of service and bucket
- sf archive-targets command added to add and manage archive targets
- archive and restore jobs now support estimates
- Updated warning for 1 million files to 1 million objects.
- sf archive and sf restore man pages updated
- Event monitor improvements
- sfdu support for multiple targets added
- daily email updated to only show errors; warnings about uid to username match are no longer reported
- API enhancements
- Added --tsv output option to duplicate_check

Upgrade Notes

- Schema version upgrade from version 46 to 47



Version 4.0.95 build 3919 (11/28/2017)

Fixes

- re:dash had redis exposed with no password. redis bound locally and password applied.
- archive job logic improvements relating to directories
- S3 compatible support--workaround for RFC7231 lack of response when requests are 0-bytes
- Largest directories for a user (recursive and non-recursive) queries don't work with UIDs, only usernames.
- Fix for schema update 44 when jobs tables already exist
- Added description to some redash queries
- Migration 35 may run in unoptimized fashion, using up excessive memory if there are large numbers of directories. The fix uses a temporary table and pre-analyzes the aggregates, instead of updating the existing table
- Unable to add volume with root inode larger than 2^{63}

New Features

- sf archive command line added
- archive feature added to Starfish beta GUI
- continued development of archive service
- sf restore now leverages archive service
- point-in-time performance improvements
- Added --show-physical-size to sfdu
- Warn when more than 1,000,000 files are in a single directory.
 - This can be modified. See /opt/starfish/etc/12-agent.ini, "dir_entry_count_warning_level" option. As with any modifications, make them in 99-local.ini
- Migrated /opt/starfish/bin/examples/jobs/sf-archive to sf archive.
 - --archive option renamed to --tar
 - --dedup, --manifest-dir, --prescan-style added
- Migrated upload-archive and upload-dedup command functionality to sf archive command.

Upgrade Notes

- Schema version upgrade from version 45 to 46
- Archive and nginx services require use of https. If your system is using http (run 'sf check-config' to determine), contact support for assistance in updating to https.



Version 4.0.94 build 3847 (11/7/2017)

Fixes

- Daily email home subject line did not include "Error" when it should have
- Fix for "QueuePool limit of size 5 overflow 10 reached" error; may cause memory leak in Postgres limited situations.

New Features

- Upload to object store includes file metadata.

Upgrade Notes

- Schema version upgrade from version 43 to 45. This also reapplies schema update for job tables, seen in 4.0.93.
 - Requires 2x size of job_result_history table and 1x size of job_result_current tables.

Version 4.0.93 build 3832 (10/31/2017)

Fixes

- Now use correct Ubuntu nginx package instead of Debian version
- sfsync-and-verify comparison reports incorrect number of file mismatches.

New Features

- List all tags on a volume feature added. Accessible both in cli (sf tag list) and API.
- Major schema update improvements, particularly related to sites updating from versions prior to 0.6.84.
 - Note that major schema updates are done 2-3 times per year, so this does not apply to minor updates, such as 4.0.90 to 4.0.93.
- Archive service infrastructure updates
 - sf restore can now restore by tag
 - tag query speed improvements
- Starfish GUI improvements:
 - Browser pane: Symlinks now show as separate icon
 - Scan pane: details window can now be hidden.
 - Browser pane: added volume to details window.
 - Dashboard added. This includes a number of existing redash reports.



Upgrade Notes

- Schema upgrade from version 41 to 43. This is an upgrade for job tables, so will take a significant amount of time for sites with large job tables.

Version 4.0.92 build 3788 (10/25/2017)

Note: May be mislabeled as 4.0.90.3788

Fixes

- Performance improvements in recursive aggregates

New Features

- Archive retrieval two pass architecture created
- Configure_redash automatically run as part of sf-redash package install/upgrade
- Web proxy implemented for upgrades
 - Nginx now proxies redash access, and provides access to API, API documentation, and Starfish GUI.
 - Nginx proxy updated from port 30000 to ports 80 and 443 for existing installs. These listen on external IP number that Starfish is configured to use.
 - Redash changed from ports 80 and 443 to port 8080, and set to listen on localhost only

Upgrade Notes

- Schema upgrade from version 40 to 41
- Access to redash is via <http://sfserver> or <https://sfserver/redash>. Access via <http://sfserver> redirects to <https://sfserver/redash>
- Access to Starfish GUI is via <https://sfserver>, for sites with beta version of Starfish GUI installed
- Access to Starfish API documentation is available at <https://sfserver/doc/api>

Version 4.0.91 build 3768 (10/17/2017)

Note: May be mislabeled as 4.0.90.3768

Fixes

- Volume larger than 8EB was not able to be added
- sf cron remove with invalid template shows success
- daily email home Vacuum Progress report error fixed



New Features

- Logging improvements for pgloader
- Archive infrastructure development
- Nginx proxy moved from port 30000 to port 443 for new installs.
- Redash changed from port 443 to port 8443 for new installs
- Added timezone support for point-in-time option
- Starfish api documentation available
 - For existing installs, use <https://sfserver:30000/doc/api>
 - For new installs, use <https://sfserver/doc/api>

Upgrade Notes

- Schema upgrade from version 39 to 40
- For new installs, access to redash is via <http://sfserver> or <https://sfserver/redash>

Version 4.0.90 build 3741 (10/9/2017)

Fixes

- Performance improvements in sf query
- redash reports no longer show reports for deleted volumes
- performance improvements in querying on files
- external monitor fixes (gpfs, lustre, isilon)
- cli crashes fixed

New Features

- Daily email home report enhancements (vacuum statistics, archive log size, postgres version and start time)
- Archive retrieval: added capability to restore directory attributes from Starfish database.
- Updates to sf query --help
- Additional performance gathering via sysstat.
- For new installs, redash is now proxied by Starfish when using https (port 443). To access via https, use <https://<starfishserver>/redash>. Access to redash via http (port 80) is unaffected.

Upgrade Notes

- Schema upgraded from version 38 to 39



- Starting with this version sysstat is now installed when sf-agent and starfish is installed or upgraded. The upgrade command documented in the Starfish Install Guide will not install sysstat, so either install it manually ("yum install sysstat") or upgrade Starfish packages individually ("yum install sf-agent sf-cli sf-core-v4 sf-nginx sf-redash starfish")
- Starfish GUI beta is now available. Contact support if you are interested in doing a beta of the Starfish GUI.
- The major version has been updated to reflect the version of Starfish.

Version 0.6.89: No release

Version 0.6.88 build 3705 (9/26/2017)

Fixes

- sf query on job with specific value fails
- sf query --ppath now works with history
- performance improvements in sf query --jobs

New Features

- Daily email home report enhanced with cpu and ram
- Install script adds option for daily email home to go to customer specified email in addition to sf-status@starfishstorage.com
- Archive service added. This will be used for Starfish archive services.
- API documentation added. Access via https://<starfishserver_IP>:30000/doc/api
 - Username is 'starfish'
 - Password is secret_key from /opt/starfish/etc/01-service.ini

Upgrade Notes

- For sites using Starfish in http mode, nginx will fail to start due to a missing ssl certificate. Contact support with assistance in switching from http to https and generating ssl certificates.

Version 0.6.87 build 3686 (9/19/2017)

Fixes



- sf query recursive aggregate "time" fixed to return max sync time.
- More gracefully fail atime/mtime that are too far in the future.
- Fix for redash report where volume capacity is 0.
- Run each database migration/schema upgrade in its own transaction, so failures in one update don't require all to be rerun

New Features

- Added restore of file attributes option to download job, since file ownership and permissions may not be identical in object storage.
- Performance enhancements when moving job results to history
- Coalesced "unable to resolve uid/gid" messages to increase readability of daily email home report.
- General work to support development and pre-release of Starfish GUI

Upgrade Notes

- None

Version 0.6.86 build 3664 (9/11/2017)

Fixes

- Improve performance of sf query --point-in-time
- redash query updates
- sf tag remove on invalid path assertion/traceback fixed.
- sfdu -B option was improperly documented as only requiring a unit; it requires both an integer and a unit (-B1G, not -BG)
- Removed license comment text shown during database migration/schema updates via pg_activity

New Features

- configure_redash has a new option ("run_now") to run pgagent daily jobs immediately. The pgagent jobs normally run at 4am and are used for longer running calculations for redash.
- historical tag retention added
- API enhancements

Upgrade Notes

- Schema upgrade from version 37 to 38.



Version 0.6.85 build 3632 (9/5/2017)

Fixes

- Detection of gpfs noatime using atime=0 setting added.
- sf query tag and tag_hum reported incorrectly

New Features

- sf cron added features via python croniter. Day (of month) and day of week is now supported.
 - Previously followed cron's default option to treat combination of day and day of week as "OR"
 - Now combination is treated as "AND" (similar to fcron), which allows for "first friday of month" (`--cron "2 0 0-6 * Fri"` will run if the day is Friday and the first 7 days of the month)
- sf query "vol_path" added as an alias for volume:full_path.
- Temp files sent to /opt/starfish/tmp instead of /tmp
- sf query --versions options added: shows all versions of single file
 - Initial version is supported for single file or small subtree only
- Retries for s3/swift requests added
- Enhanced reporting of database migrations (schema updates) to postgres logs

Upgrade Notes

- None

Version 0.6.84 build 3598 (8/28/2017)

Fixes

- sfsync-and-verify.sh updated to allow target directory to be empty on first run.
- Added "--allow-empty-directory" option to sf volume add to allow addition of empty volume.
- sf query --mode changed from decimal to octet, to better match standard usage.
- pgloader failure due to dates after year 30000 resolved.
- Various --help updates
- Various re:dash dashboard/query updates
- Postgres autovacuum optimization

New Features

- Number of blocks added to sf query --aggrs.



- Min/Max atime, mtime and ctime added to recursive aggregates
- sf query case insensitive name search option added. Use --name for case sensitive and --iname for case insensitive.
- Improve sf query --name performance
- sf tag bulk-add "from-file" format changed. Previously one specified a volume and the from-file had a list of full_paths for the volume. The new format includes both volume and full_path in the from-file (sfvol:path/to/file).
 - Using sf query with the --format "vol_path" output option will produce the correct input for the sf tag bulk-add --from-file option.
 - If including multiple volumes in the --from-file, make sure sort/order the files by volume, since Starfish batches them by volume.

Upgrade Notes

- Schema upgrade from version 33 to 37. See "Upgrade process for blocking schema updates" below for upgrade notes. Note that this is a major schema update that may take a significant amount of time.
 - Requires 3x size of dir_history table, and ½ size of dir_current+file_current.
- If you are not running redash 1.03, contact support for help upgrading to version 1.03.

Version 0.6.83: No release

Version 0.6.82: No release

Version 0.6.81 build 3542 (8/14/2017)

Note: This is an updated release of version 0.6.81.3516

Fixes

- Sf collect-logs did not properly gather job logs, only scan logs.
- Redash installation: failure to add redash to init.d when /opt/starfish is a symbolic link or mount point.
- Minor fixes in redash reports

New Features

- Scan of empty volume
- When running "sf volume add", it notes that the volume will be scheduled for a 9pm daily diff scan.
- Daily email now shows errors for nfs v4 mounts where filenames do not display.

Upgrade Notes



- None

Version 0.6.81 build 3516 (7/31/2017)

Fixes

- Changed --timestamp option in sf query to --point-in-time
- Redash: tag pivot table added mtime
- Redash query optimization
- Noatime detection on automounted filesystems fails.

New Features

- Added --point-in-time option to sfdu
- Expanded--point-in-time option to allow date and time, instead of date only.
- Pgloader parallelization implemented
- Added sf query to show jobs that exist. Example:
 - sf query sfvol1: --jobs.hash:exist

Upgrade Notes

- Schema upgrade from version 29 to 33. See "Upgrade process for blocking schema updates" below for upgrade notes.

Version 0.6.80 build 3473 (7/19/2017)

Fixes

- re:dash administrator user renamed to "Starfish Administrator". No changes to login name or password.
- sf-meta example job updated to use jq, so Starfish metadata will not be added as an array inside an element. This allows Starfish to query for individual items added via sf-meta job.
- Increased cron service timeout from 10 to 30 seconds.

New Features

- Dashboards reorganized into Volume, User, Tag, Tier and Starfish dashboards.
- Additional queries added to dashboards



Upgrade Notes

- Suggest upgrading redash to version 1.0.3. Contact support for assistance.

Version 0.6.79 build 3444 (7/12/2017)

Fixes

- Sfsync-and-verify compare phase optimized for Starfish v4
- Hash job from duplicate_check hangs
- Duplicate_check compare phase optimized for Starfish v4
- sf check-config displays incorrect agent address

New Features

- Alias stop and kill in sf job and sf scan
- Add rec_aggrs to query/filter in sf query.
- Timestamp option added to sf query. This allows querying of historical scans.

Upgrade Notes

- None

Version 0.6.78 build 3424 (7/5/2017)

Fixes

- Update example jobs to put temporary files in /opt/starfish/tmp
- Dispatcher fails to start when more than 1 job is to be resumed
- Add variable to disable/reduce sysstat messages in daily report

New Features

- Add starfish-nginx package. This will proxy all Starfish services in the future.
- Add sf-vol-rename script to /opt/starfish/bin/tools.
- Add numeric job id
- Add sf-job-rename.sh script to tools directory. This is used to rename mount point or agent for a Starfish volume.



Upgrade Notes

- None

Version 0.6.77 no release

Version 0.6.76 build 3383(6/23/2017)

Fixes

- Pgloader loops on broken chunk
- Dispatcher fails to start when more than 1 job is to be resumed

New Features

- Minor updates to new dashboards and queries
- Remove OTHER from computed tags
- Add default cost to volumes on creation
- Updated re:dash version to 1.0.3. Previous version (.11) still supported. See upgrade notes section of this document for changes in `configure_redash` usage.

Upgrade Notes

- Redash version upgraded to 1.0.3. Contact support for the `upgrade_redash.sh` script to automate removal of old version and installation of new version. Note that the default login for re:dash changes from starfish (password Starfish123) to `starfish@starfishstorage.com` (password Starfish123)

Version 0.6.75 build 3354 (6/14/2017)

Fixes

- Augeaus parse error during installation

New Features

- Added new reports to re:dash, including:
 - Largest Directories for a User (recursive and non-recursive reports)
 - Tagging reports: Overview, size by tagged directory, and largest directories within tagged directories.
 - Churn and growth reports



- User size change rate
- Estimate Duplicate report (simple size comparison only)
- Volume growth rate over time
- User aging

Upgrade Notes

- None

Version 0.6.74 build 3325 (6/7/2017)

Fixes

- Removed extraneous queries added to Starfish dashboard in re:dash, reverting to version 0.6.72 view.
- Removed “-f csv” option. This option was available in Starfish version 3, and was superceded with --csv in Starfish version 4.
- Documentation updated to include additional package requirements for CIFS scanning and AD integration.
- Now able to remove job that is “starting” or “resuming”

New Features

- Added re:dash extension report by number of files to Starfish dashboard
- Installer updates: install redash in /opt/starfish/redash; set umask to 0022 prior to installing redash, if necessary;

Upgrade Notes

- Starting with this version, `configure_redash` will remove any Starfish installed dashboards or queries. It will not remove any dashboards or queries that Starfish did not create. Make sure to fork any Starfish queries you modified prior to running `configure_redash`.

Version 0.6.73 build 3279 (5/30/2017)

Fixes

- Update to support S3 signature version 4.

New Features



- New re:dash report: churn report
- Added logrotate for pgagent.log, since pgagent.log does not integrate timestamps in currently released Postgres versions.

Upgrade Notes

- None

Version 0.6.72 build 3239 (5/23/2017)

Fixes

- Timeout increased for large scans that failed during preparing phase.

New Features

- Starfish services updated to use https
- New re:dash reports
- Enhancements to daily email home report (report_stats.sh)
- Added logrotate for pgagent.log, since pgagent.log does not integrate timestamps in currently released Postgres versions.

Upgrade Notes

- None

Version 0.6.71 build 3235 (5/19/2017)

Fixes

- sf collect-logs updated to ignore log files deleted during collection.
- Supervisord configuration updated to start pgagent regardless of Postgres version. (CentOS only)
- Slow job stopping resolved

New Features

- Postgres backup option addition (cont'd)
- Starfish licensing addition (cont'd)
- Enhancements to daily email home report (report_stats.sh)

Upgrade Notes

- Daily email home will be disabled after upgrade. To activate, edit /etc/sysconfig/starfish and set "SENDING_DIAGNOSTIC_REPORTS_ENABLED" to "true". Note that email must be sent as the



root user, or postgres logs will not be included; new installs now use starfish user's crontab with "run as root" option for report_stats.sh scheduling.

- Existing sites should automatically pull license from repository. If sf check-config has any license issues, contact Starfish support for a license file.

Version 0.6.70 build 3188 (5/5/2017)

Fixes

- Remote agent daemon does not start.

New Features

- New re:dash report: simple duplicate file report.
- Added -H alias for --no-headers in sf query.
- Enhancements to daily email home report (report_stats.sh)
- Postgres backup option added. (Run current installer to activate)
- Starfish licensing added

Upgrade Notes

- Final version of Starfish v3 (mongo).



Upgrade Process

Upgrading Starfish is done via yum or apt-get, and is documented in the Install Guide. The following is the process:

1. Stop jobs and scans:
 - a. Run "sf scan pending" and "sf job pending" to determine if any scans or jobs are running.
 - b. Stop any running scans by typing "sf scan kill -a"
 - c. Stop any running jobs by typing "sf scan kill <JOB_ID>"
 - d. Rerun "sf scan pending" and "sf job pending" to determine if any scans or jobs are running.
2. Run "yum upgrade starfish" to update the server, and "yum upgrade sf-agent sf-core-v4 sf-cli" to upgrade an agent, on RHEL/CentOS.
 - a. "yum upgrade" is also supported, but will also upgrade non-Starfish packages.
3. After upgrading, run "sf check-config" to verify that all services are up and running.
 - a. If a schema upgrade is in process, several services will indicate that ("db migration in progress")
 - b. If the agent is not running, start it ("service sf-agent start"), then run "sf check-config" to verify it is up.
4. If you changed the password for starfish@starfishstorage.com in redash, run
"/opt/starfish/bin/configure_redash default --email starfish@starfishstorage.com --password <PASSWORD>" to update redash with current set of dashboards and queries. Note that this should only be run after "sf check-config" returns with a healthy status.

Notes:

1. All Starfish nodes (Starfish server, agents and cli systems) must be updated at the same time.
2. Configure_redash removes and replaces all standard dashboards and queries, but leaves alone any non-standard ones. If you have customized redash queries or dashboards, fork them off prior to running configure_redash.



Other Notes

Logins for Starfish are as follows:

1. Starfish CLI: Users in "starfish" group have full access, as does root. Users in starfish-ro group have access as defined in `/etc/sudoers.d/starfish-ro`: `sf query`, `scan list`, `job list`, and similar.
2. redash: Default login is `starfish@starfishstorage.com`, password `Starfish123`. If you have modified the password and lose it, you can either reinstall redash or install the redash cli tools to recover it. Unless you have custom dashboards or queries, the former is significantly easier. Support can help with both.
3. Starfish GUI: Username is `starfish`, password is the `secret_key` from `/opt/starfish/etc/01-service.ini`