Archival Storage

Content

Further information

Technical infrastructure

Preservation of data integrity as part of the process routines

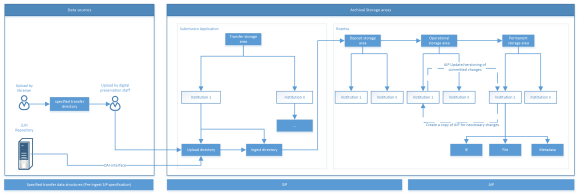
Specifications for archival information packages (AIPs)

Oracle Solaris ZFS Administration Guide

Storage areas

Archival storage is divided into different storage sections for the various statuses in the object's lifecycle. The storage sections contain separate storage areas for each institution in the alliance of three German national specialist libraries; objects and their corresponding metadata are deposited in these storage areas. Each institution is only able to view its own objects.

Archival storage is divided into different storage sections for the various statuses in the object's lifecycle.



- Transfer storage area with the subsections upload and ingest:
  - Upload: This directory is used to upload objects that are ready for ingest. It is the source directory for the submission application.
  - Ingest: This directory is the target directory for Rosetta-compliant SIPs created by the submission application.
- Deposit storage area: This directory is used to copy SIPs from the ingest directory and to enrich them with metadata from the ingest process. Each SIP is designated a unique ID.
- Operational storage area: This area is used to copy ingested SIPs from the deposit storage area that were ingested with a manual workflow and further processed; for example, access copies are added. SIPs that have been redirected to the technical analysis area, and the temporary copy created during the update of an AIP are also deposited in this area.
- Permanent archival storage with the subsections file, IE and metadata: TIB uses the concept of the logical AIP, and saves the files, the METS and the catalogue metadata in physically separate places. Before processing, a copy of the AIP is created and moved to the operational storage area. If modifications are confirmed, the AIP is versioned and the new version is deposited next to the original AIP.
  - File: This area contains all files within a version of an AIP.
  - IE: This area contains one METS file for each version of an AIP. The METS files contain references to the files and the XML file with the catalogue metadata that belong to the relevant AIP.
  - Metadata: This area contains an XML file with the catalogue metadata for each identifier.

Storage rules configured in Rosetta define where which objects are stored. The storage rules can be adapted and documented in a consortial configuration description, which is adopted by the partner libraries on an annual basis.

Preservation of the completeness and intactness of archive packages

Storage media

Two independent NAS systems are available for the digital archive's archival storage. These are managed using ZFS file systems and operated as RAID-Z3 systems. Each RAID-Z3 system contains several logical hard disk arrays, each consisting of several physical drives. NAS 1 is a high availability cluster, and is used as the production system; NAS 2 is used for disaster recovery. The currently available storage can be extended up to approx. 3.8 PB (gross).

ZFS supports data integrity assurance by means of integrated checksum procedures and the RAID-Z3 system's self-healing features. ZFS works with Copy on Write (COW), so that the file system remains consistent even after power cuts and system crashes. This feature is particularly important in the case of data mirroring from NAS 1 to NAS 2 (see Redundancy).

RAID-Z3 is a hard disk array with triple parity.

Redundancy

Files are distributed in the RAID-Z3 array and are stored so that files are recoverable to a certain degree.

ZFS Send and ZFS Receive are used to replicate an incremental snapshot to the second NAS system at a set time each day. Checksums ensure consistency of the data during transfer.

Monitoring and refreshing

The NAS systems are geographically separated from each other in separate, locked server racks in the TIB data center (NAS1) and in the Leibniz University IT Services (LUIS) data center (NAS 2).

The server racks are equipped with temperature monitoring and a gas extinguishing system for each shelf. Each storage system has a reporting tool that enables storage capacity, the condition of the hard disks, and tasks such as replication to be monitored automatically. In the event of a hard disk failure or defect, the system automatically triggers a message to the administrator.

Access to the data centre is secured by an electronic access system and a burglar alarm system, and is restricted to a few members of staff only. The data centre is equipped with a fire and smoke detection system, as well as an independent emergency power supply that enables the proper shutdown of the servers in the event of a power cut.

Media migration

TIB has a service contract with a service provider concerning the replacement of defective hardware. Monitoring software automatically notifies the service provider when defective hardware needs to be replaced. The service provider then supplies replacement hardware, which is installed without disrupting IT operations. The resilvering ZFS function restores data integrity following a media failure.

Integrity assurance

One checksum is generated and saved per block. A checksum is generated with every read access and every data transfer, and is compared with the saved checksum. If the checksums do not match, the corrupted block from distributed storage is restored in the RAID-Z3 system.

If the checksums do not match following replication from one NAS system to the other, the transmission is aborted and reported.

Where necessary, a checksum comparison (ZFS scrubbing) may be run for all blocks on the file server to check the data integrity of the whole storage system. This check is carried out automatically every 90 days.

A complete snapshot is generated for each NAS system daily and before maintenance work. This enables changes to a system to be viewed, traced and, in case of need, reimported retrospectively for seven days. Snapshots from the past 30 days are also stored.

The digital preservation software Rosetta also features mechanisms for preserving data integrity.

Recovery

The self-healing feature of the chosen ZFS system may compensate for the failure of up to three hard disks from a ZFS stripe. Recoveries within a disk pool can be viewed in the monitoring software.

If the RAID-Z3 system is no longer able to perform the recovery, the Systems Administrator initiates disaster recovery from the replication partner. If NAS 1 has sustained irreparable damage, the files, including the file system from NAS 2, can be fully replicated to another system, which then assumes the function of the production system. Checksums ensure the consistency of the data during transfer. The Systems Administrator must initiate and monitor this process.