



Chapter 3: Replica Management

Contents

- Replication rules
 - Concept of requesting replication for current data
- Subscriptions
 - Concept of requesting replication for future data
- Replica Deletion

Replica Management

- Replica management in Rucio is based on **replication rules**
- A replication rule defines the minimal number of replicas to be kept on a set of RSEs
 - e.g.: *2 replicas of file user.jdoe:file_001 on any TAPE system*
- Declarative data management instead of imperative data management
 - “Two copies of X on TAPE” vs “Copy of X on TAPESYS_4 + Copy of X on TAPESYS_9”
 - Easier to use, optimize storage space, minimize number of transfers
- Multiple ownership of data
 - Rules not only invoke transfers, but also protect data from deletion

Replication rules I

- Replication rule is created by an account for a data identifier (file, dataset, container)
 - Rule affects all files in a dataset or container
 - Rules create transfer requests, if they cannot be satisfied with existing replicas
- Rules are enforced permanently (until they are deleted)
 - Files added/removed to dataset, dataset added/removed to container
- Arbitrary amount of rules can be defined for the same data identifier

RSE Expressions

- RSE Expressions are used to describe a set of RSEs
 - E.g.: All DISK RSEs in Germany: `country=de&type=disk`
 - Set-complete language
 - All RSE Attributes can be used
 - Primitives:
 - Union `|:country=us|country=fr`
 - Intersections `&:country=us&tier=1`
 - Complements `\:country=de\type=tape`
 - Comparison `> <:freespace>150`
 - Parentheses can be used to define order of operations
 - `CERN_EOSDISK|(country=de&type=disk)`

Replication rules II

- Required parameters for a rule are: number of copies, did, RSE Expression
 - E.g. 3 copies of data:dataset1 on (country=de|country=fr) &type=disk
- Rules get enforced continuously, but are not re-interpreted
 - Order of creation matters:
 - 1 copy on RSEa; 1 copy on RSEa|RSEb → 1 physical replica on RSEa
 - 1 copy on RSEa|RSEb; 1 copy on RSEa → 1 physical replica each on RSEa and RSEb
- Rules can have a lifetime, after which the replicas become eligible for deletion
- Grouping options (for rules on dataset/container) for rules
 - DATASET (When rule is on a container): Each dataset is distributed to the same RSE
 - NONE: All files of the dataset/container are distributed randomly
 - ALL: All files of the dataset/container are distributed to the same RSE
- Different notification modes to notify creator as well as external applications

RSE Selection

- Primary objective is to **minimize the creation of transfers**; This has precedence over everything else
- If there are overlaps between a new rule and existing replicas, the rule will try to re-use these replicas in a way to create the least amount of transfers
- If there are no overlaps, RSEs are selected randomly, unless the weighting option of the rule is used
- Permissions and Quota are enforced
 - No Quota → RSE will be excluded
 - RSE write blacklisted → RSE will be excluded

Quota

- Quota (account limit) = amount of bytes available to a user on an RSE
 - Users pay quota for creation of rule
 - Site administrators are able to specify quota
- Example:
 - 3 users have a rule each on file data:file1 of 100MB on RSEa
 - All 3 are paying 100MB of quota on RSEa; not only the first user responsible for creating the physical replica

Rule approval

- Users can create rules arbitrarily within their quota
- If users have no quota or not enough quota they can ask for manual approval
 - Sent to a list of “approvers” set per RSE
 - The approvers can approve/deny the rule
- Manual approval can be disabled for an RSE
- Automatic approval can be enabled for an RSE
 - E.g. all rules smaller than 500Gb requested for approval are automatically approved
 - Limit is settable per RSE

Subscriptions: Automatic replication of new data

- Replication rules cannot be created for future data, as they require an existing data identifier
- Subscriptions create replication rules on newly created data
 - Each subscription consists of a filter of metadata to match
 - And a list of replication rules to apply
- All newly created data identifiers are matched against all subscriptions
 - If a match is found, the rules are applied
- Example
 - All newly created RAW detector data should be replicated to 2 tier-1 disks and 1 tier-1 tape system

Replica deletion

- Replicas not covered by a rule are eligible for deletion
- Actual deletion depends on the deletion policy setting of the RSE
 - Standard: Only delete replicas when space is actually needed on the RSE
 - Greedy: Cleanup space as soon as possible → Replicas are deleted once the rule is removed
- Standard deletion mode keeps replicas indefinitely, unless space is needed on the RSE
 - Threshold can be configured
 - Order of deletion is based on LRU replica

Questions?

If you have a question but don't get the chance to ask it directly during the session, you can do it here: <https://goo.gl/BdSGoC>