Relativity®

Workflow Solutions for Very Large Workspaces

December 8, 2021 | Version Server 2021

Table of Contents

1 Overview

Very large Relativity workspaces (VLRWs) require a great deal of time and effort to maintain. As a result, it's important to develop a plan to accommodate the workflow before your workspace reaches VLRW status. Making workflow changes after this point is inefficient and can be very time consuming.

Use the following best practices to plan your VLRW workflow and improve overall performance. These guidelines can also be a helpful starting point for educating Relativity users. This guide also outlines the best practices for setting up and working in VLRWs.

2 Importing data

Importing large amounts of data can be time consuming.

Consider the following recommendations:

- Break the load file into smaller document counts.

- Use multiple computers and instances to load files.

- Load the control numbers and folder paths into the workspace to create the records in SQL.

- Load text as external text files.

- Break each subgroup load file into reasonable sizes, such as 250,000 records per load file.

- Be aware of the kinds of data that you import. Importing the extracted text from large binary files, such as WAV, MPEG, and Access database file types cause excessive load. Don't attempt to load these file types.

Note: To more easily verify data and recognize errors, load in smaller batches of data and verify each batch as you go.

3 Fields

Fields are not endless commodities. Too many or overly bloated fields can dramatically impact database performance. Use the following guidelines to improve field efficiency:

- Keep fixed-length text fields to the minimum size needed.

- Set fixed-length text fields to no more than 400 characters.

- Store choice data in single-choice or multiple-choice fields if possible.

- Use separate objects to store repetitive content when possible.

- Set up fields as Unicode in advance because the system must re-index a field if the change is made following data load.

- Only include fields in the text index when it's necessary to do so.

4 Folders

You can use folders to organize documents to reflect their original storage. However, in a large workspace, the use of folders can cause performance issues. Folders perform searches in the background to display documents. If searches begin to take longer as more documents are loaded, consider the following options:

- Carefully place an index on the custom fields to assist in pulling up documents as the folder size grows.

- Combine folders and then use custom views to achieve similar organization.

- Create a multiple-choice field and populate the field with the folder path. This will create an entry in the field tree to allow for organizing the documents by the original folder path and users can folder documents in multiple locations.

5 Index management

Optimized indexing requires some knowledge of your data. Scrubbing your data before indexing saves time when creating an index and returning search results.

Consider the following when creating an index:

- Remove file types that have no searchable content, such as system or program files.

- Use a separate index for searching database files and large Excel files.

  Note: Even if your database has only a small number of these files, creating an index without them improves searching speed.

- Set up multiple dtSearch indexes, including one with a smaller document set based on one or more of the following criteria:
    - Date ranges
    - Custodians
    - Text size (extracted or OCR text)
        - Small (< 2 MB)
        - Medium (> 2 MB and < 10 MB)
        - Large (> 10 MB and < 25 MB)
        - Very large (> 25 MB)

Communicate with your team to create a search strategy for the case. Some cases have distinct words or terms that might warrant changing the default settings of an index.

  Note: Editing these settings may affect search results.

- Remove numbers from the dtSearch index alphabet file if you're only searching for words—this reduces the size of the index and disables numeric range searching.

  Note: Be sure to communicate any changes to alphabet files. Searching against multiple dtSearch indexes that use different alphabet files can result in different results, even when running the same search on identical content.

- Set the dtSearch index to recognize and/or ignore words, characters, and digits as necessary. While these settings don't necessarily impact performance, applying them ahead of time avoids having to rebuild the index later.
  For example, if a company name appears many times throughout a document set and you don't intend to search for it, add this name to the noise words list. Configuring these settings before building a large index prevents you from having to rebuild the entire index later to include these types of characters.

- Enable dtSearch indexes to automatically recognize dates, email addresses, and credit card numbers only when necessary. Enabling this setting increases build time.

- Consider using a pair of dtSearch indexes when adding new data. You can update one index in the background and then swap out the outdated index with the current one.

6 Layouts

You can use layouts, along with views, to improve workflow efficiency. Identify the type of information each reviewer group needs to code documents. (e.g., a group may be working on privilege logs, prepping for depositions, etc.). You can then use group security permissions to adjust layout visibility as necessary.

When planning layouts, think about the overall life-cycle of a document. For example, a review workflow may include the following:

- First pass review
- Quality assurance
- Production
- Privilege review
- Deposition prep
- Trial exhibits

Some users have many layouts that they need during the course of a document review. You can use separators (-----------) to help organize layouts and build the workflow.

Issue coding layouts can get long and cumbersome over time, requiring users to scroll to see all available choices. To improve the layout's usability, change the layout field display from check box list to pop-up picker. This declutters the layout space by hiding the choices and presenting them only when necessary. Users can apply filtering to pop-up picker views to find choices.

7 Mass operations

Mass operations temporarily lock the document table while executing. Depending on the number of records and users in the system, the table may lock for an extended period of time and frustrate users trying to perform standard edits. If necessary, carry out mass operations at night or at an off-peak time.

8 Persistent highlight sets

Persistent highlight sets provide a valuable way to identify terms within the document viewer. Although the size of a workspace doesn't affect how persistent highlighting works, use these guidelines to improve usability in large workspaces:

- Enter multiple terms on separate lines.

- Enter terms exactly as they appear in the document. Don't use quotation marks or connectors.

- If you enter variations of a term or phrase and the variations include multiple words, list the multi-term variations first. The regular expressions that persistent highlighting uses look for and find a term, and then they move to the next term in the set. For example, you should list the terms United, United States, and United States of America, in the following order:
    - United States of America
    - United States
    - United

- Don't use special characters, quotation marks, or other punctuation.

- Don't use dtSearch syntax, including operators such as "AND" and "OR."

- Identify and remove terms with large hit counts.

- List variations of a term first and the root term last.

- Use Highlight Fields and Search Terms Reports to generate persistent highlight sets.

9 Exporting data

Exporting large productions takes a great deal of time. Create saved searches to divide the production into roughly equal amounts, such as approximately 250,000 pages each. On multiple machines, export each saved search. Use the production images as the default.

For example, the following process exports 250,000 records with approximately 2.5 million pages to a network share folder:

1. Create saved searches of the production so that each has approximately 500,000 pages.

2. Label each volume in sequential numeric order.

3. Modify each image load file to show the top folder as the production volume.

4. Combine each load file into one complete load file.

5. Export each saved search of images using a single machine.

6. Export any native files on a single machine, selecting the Beg Bates field only.

7. After the exports, create a fixed-length text field called "Prod Native Path".

8. Use the Relativity Desktop Client to overlay the exported load file from step 6 onto the Prod Native Path field.

9. Export the text files for each record using a single machine, selecting the Beg Bates field only.

10. After exporting the text files, create a fixed-length text field called "Prod Text Path".

11. Use the Relativity Desktop Client to overlay the exported load file from step 9 onto the Prod Text Path field.

12. Export the metadata for all the records, after loading the information for Prod Native Path and Prod Text Path.

Depending on the speed of your environment, this process may assist in the export process.

System admins have all necessary permissions to perform the following script-related actions:

- View

- Run

- Preview (locked and unlocked scripts)

- Create/Write

- Edit

- Link Import Applications (See the Applications Guide)

Users should not frequently run custom scripts that can have a negative impact on the system, including through SSMS access. Avoid using SSMS and the Admin Script functionality as much as possible until actions are audited and certain Relativity controls, including timeout values, are in place.

To prevent scripts from negatively impacting your environment:

- Limit admin script access for a given workspace to one or two people.

- Assign an individual to review the impact of custom script executions on the system.

Once you've identified the scripts safe for execution, you can make them available to users through the workspace tabs.

10 Reviewer statistics script

Reviewer Statistics is a popular script in the Relativity script library that reports on the number of documents reviewed over a certain period of time. It can take a while to complete in workspaces with large audit record tables. Instead of trying to run this report regularly during review, we suggest that you schedule it to run each night after maintenance has completed. You can then have the results emailed to one or more recipients.

11 Searching

Executing searches can be very resource intensive. Follow these guidelines to reduce the resources used for searching.

- Don't use the "is like" search operator on Fixed Length and Long Text fields. Using "is like" runs a resource-intensive bit-by-bit search rather than using the index.

- Avoid using multiple layers of nesting applied in a search.

- Don't use wildcards in the front or in the middle of terms. Instead use the dictionary to find multiple forms of words and paste all of them into the search box.

- Avoid searching on a number of unnecessary search terms. Instead, use the dtSearch Dictionary's fuzzy and stemming searches identify the best words to search.

- Avoid searching the entire dataset. Instead limit the search to subsets of data, such as certain date ranges or custodians.

- Use filters as an alternative approach to searching.

12 Search terms reports

Search terms reports (STRs) simplify the process of identifying documents that contain a specific group of keywords. Instead of running complicated queries, you can use the search terms report to enter a list of terms or phrases and then generate a report listing their frequencies within a set of documents. As workspaces grow in size, a search terms report takes longer to run if the individual string is too complicated.

In large workspaces, avoid using nested proximity searches or wildcards in search terms reports. Nested proximity searches run slowly in large dtSearch indexes because the search string takes longer to search the index. Using wildcards before a term, after a term, or on both sides of a term, causes the search terms report to take much more time to complete. Instead of using wildcards, use the dtSearch Dictionary to identify variations of a term.

Combining wildcards and nested proximity searching may create overly complicated searches. This adds a significant amount of time to running a query and sometimes prevents it from completing. For example, (((Term1* or Term2*) w/20 Term3*) and Term4*) and (Term5* w/20 Term6*) is a complicated query.

13 Security permissions

Large workspaces usually require multiple security groups. You should organize documents and define security groups to assist with review workflow. Start with a baseline security group for each main role.

For example, you may need to create a baseline group for each of the following roles:

- System admin
- Operations admin
- Operations technician
- Project manager
- Project specialist
- Case admin
- Case review
- Case technician
- Contract reviewer
- Experts

Set security permissions from the baseline, giving each user group incremental security rights as necessary. For example, three different user groups may need permissions for the following tasks:

- Contract review
- Contract review – QA – Access to QA layouts, fields, choices
- Contract review – Privilege – Access to Privilege Log layouts, fields, choices
- Contract review – Dep Prep – Access to Deposition Prep layouts, fields, choices, mass actions

14 Snapshot auditing

Enabling Snapshot Auditing On Delete increases the database size.

15 Views

When it comes to the reviewer interface, focus on workflow. Create views to filter document collections to necessary lists. Using views, you can manage the types of documents that are presented to a group. Use group security permissions to turn views toggle view visibility as necessary.

Analyze each group participating in the review, and map out its exact needs. For example, the First Pass group only needs to review batches. The Second Pass group needs to both review and quality check documents, and the Experts group only needs to see Production documents. Implementing a plan that coordinates views for review groups improves workspace management efficiency.

In addition, consider the following best practices for views:

- Add only the fields necessary for the review task to a view.

- Avoid adding Long Text fields to views.

- Using nested saved searches as view conditions slows down loading.

Proprietary Rights

This documentation ("**Documentation**") and the software to which it relates ("**Software**") belongs to Relativity ODA LLC and/or Relativity's third party software vendors. Relativity grants written license agreements which contain restrictions. All parties accessing the Documentation or Software must: respect proprietary rights of Relativity and third parties; comply with your organization's license agreement, including but not limited to license restrictions on use, copying, modifications, reverse engineering, and derivative products; and refrain from any misuse or misappropriation of this Documentation or Software in whole or in part. The Software and Documentation is protected by the **Copyright Act of 1976**, as amended, and the Software code is protected by the **Illinois Trade Secrets Act**. Violations can involve substantial civil liabilities, exemplary damages, and criminal penalties, including fines and possible imprisonment.