

Blackboard

About "Course Size" and Reducing Storage Utilization on Blackboard Learn

Date Published: Sep 27, 2020 **Category:** Product:Content_Management_Learn;

Version:Learn_9_1_Q4_2019,Learn_9_1_Q2_2017_3200_0_0,Learn_9_1_Q4_2017,Learn_9_1_Q2_2018,Learn_9_1_Q4_2018,Learn_9_1_Q2_2019,Learn_9_1_Q4_2016_3100_

Article No.: 000057526

Product: Blackboard Learn

Release: 9.1;SaaS

Introduction: From time to time, Blackboard Learn clients across all hosting options may want to delete old courses and content to reduce space usage. The Xythos Web File Server application, which forms the content system of Learn, has many powerful features, including features to reduce disk utilization from the outset, but they may cause a lower than expected amount of space to be recovered after mass deletions.

There are two major prongs to this:

- Deduplication: pre-optimizes storage and means little might be available for recovery.
- Deferred Delete: prevents any changes from being in effect for 32 (SaaS) or 45 days.

Functionality:

Deduplication

The Blackboard Xythos web file server application that manages content files, other than legacy storage, abstracts *metadata*, (information about the file such as the name, path, permissions, etc) from the *data* of the file itself.

In other words, if 500 people upload 500 copies to 500 courses of the same ½ megabyte clip-art file, the expected total disk consumption is still ½ megabyte, not 250 megabytes. Xythos 'hashes' the file's data and realizes it is already stored. The new data is discarded and Xythos creates additional metadata records that point to the existing data. Therefore, no file space will ever be recovered until every last "pointer" to the data itself is removed. In the clip-art scenario above, it is necessary to delete 500 courses to recover the space the clip art occupies.

What this means in practice, is that most disk savings have already taken place and it is common for 'surprisingly' little space to be recovered even when large amounts of courses are deleted. The Xythos application pre-optimized the utilization and has already saved those "costs" from the very beginning. Sites never "incurred" the storage that was saved by deduplication, therefore inversely there is not "future savings" left on the vine.

Deferred Delete

Xythos keeps count of how many "files" refer to a given piece of data, represented as a physical back-end storage file. Once the 'count' of references to a given piece of data drops to zero, the data is marked as "eligible" for deletion. Eventually, it will be cleaned up. A subsystem called the XythosStorageThread examines the system periodically. Among its functions is to clean up back-end physical storage data files that are "eligible for deletion" and have been so for longer than the deferral interval. These are then deleted permanently. On Managed and Self-Hosting the deferred deletion interval is 45 days. On SaaS it was initially 45 days but changed to 32 days in September, 2020.

- [Changes to Storage Retention Policy in Learn SaaS Sites](#)

If someone adds a new copy of the file, that is byte-for-byte identical to data that is in the deferred delete status, it will immediately be taken out of this state and the reference count will increase by one.

On SaaS the files in the deferred deletion status reflect as "Deleted Files" in the [SaaS Insite Report](#).

A Practical Example

The majority of courses will be copied from other courses, as most courses are offered more than once. Accordingly, if ENG-101 is copied from a template in the spring, then copied from spring to summer, from summer to fall to intersession and back to spring anew in the next year, all those files that are part of the main content of the course may incur no further storage use because they are byte-for-byte identical.

New database records are created to indicate the paths, permissions, access and so-forth, but the actual physical back-end storage files for their data, which present the vast majority of the 'utilization' of space, is not stored again and again. The only files that are going to incur additional storage are new files (never-before existing), such as new materials that the instructors upload, students' submissions to assessments, attachments to discussion board postings and things like this.

Limitation of the GUI Report

It is not possible to take any of these effects into account from the disk GUI report (course size report). Therefore, common questions such as "how large is this course?" and "how many megabytes of space does this term take up?" cannot be answered easily. Course size is not a well-defined feature of Learn. The GUI report will show deletions immediately, and will count data stored-in-fact-but-once as many times as there are records in that course. In other words, the GUI report thinks that the clipart from the first example is worth 250 megabytes.

The output of commands such as du for Self-Hosting and the Managed-Hosting Insight report will always diverge from the GUI Disk report because it takes into account the deduplication and deferred delete interval. The UNIX system always shows the actual, physical storage without complications. There is no good theoretical way to present a graph or other indication that shows the deduplication factor, so it is best to consider both reports.

Exception - "Legacy Storage"

Legacy storage is as it is named and predates the acquisition and incorporation of the Xythos company and product into Learn. Legacy storage may consist of older course messages and course message attachments in courses created prior to Learn 3300.0.0, as well as SCORMs from prior to Learn 3500.3.0. There can also be very old course file attachments. However, in a recently-created course in a modern site, there will be little or no legacy storage. A few older content types such as Self & Peer Assessments, and Glossaries still create files, but these do not commonly increase to significant size.

Prior to Learn SaaS 3800.6.0, when a course was deleted, the 'Protected' (Internal) files described below were copied into the Legacy recycle bin.

Legacy SCORMs and Course message attachments are a common source of large consumption. Migrating course messages can reduce this utilization but is rather intensive. Because Legacy storage does not deduplicate at all, and is not eligible for deferred deletion, targeting courses which have a lot of legacy storage is an effective strategy for reducing overall storage consumption.

What are 'Protected' (Internal) Files?

In addition to course archive and export files, there are some other attachment types that are stored in the "/internal" folder. The disk GUI report refers to this as the 'protected' storage. Internal storage is mostly student submissions, attachments to discussion boards and modern (post 3300.0.0) course messages as well as some other miscellaneous files that are usually unique to a specific course. From Learn 3500.3.0 and later, SCORM packages began to be stored here.

To reduce the consumption of internal storage, please delete old Archive and Export files from each course, and check for practices such as 'icebreaker' assignments that involve recording a video of one self speaking for a few minutes. With HD video commonly available on phones now, these submissions and discussion board posts can incur significant storage cost. Upload those kinds of videos to Youtube, Kaltura, Panopto, or similar services.

Technologies:

Recovering Space Quickly

If it is needed to recover as much space as possible now, then it will be necessary to lower the deferred delete interval. This can be set to as low as "1" day (do not set it to zero days). Please contact blackboard support. This option is unavailable on SaaS.

Because Legacy content is not deduplicated, and does not have deferred deletes, targeting courses with large amounts of legacy storage can be effective, as can migrating legacy content to the Content System.

- [Move Legacy Content to Content System \(Self-hosted Windows/Linux\)](#) - Article #43037

Targeting courses with a great deal of internal ("protected") storage can also be effective. With the principal exception of usage by modern SCORM packages, these files tend to be unique, so the deduplication effect is not as pronounced. However, they are still subject to deferred deletes.

Do not Attempt to Delete Files from the Physical Storage Folders on the Content Share

There is no approved means to do this. In addition, a watchdog process proactively spot-checks random selections of data files. If it finds many of them to be missing compared to database records, it marks the storage location as being "not mounted properly." This takes it offline, causing an outage, until the administrator corrects this problem, such as by restoring the storage from backup. In this condition, much of the Learn system will then stop working until it is resolved.

Planning for Mass Deletion of old Courses on Managed Hosting and SaaS

If it is needed to to archive-and-purge old courses, please create a new case with a list or criterion. Afterwards, when exportation or archiving is complete, Blackboard Client Support will provide instructions to purge the courses. Please note that the 'archive' files created from the mass archiving request are not going to be stored in the site at all; they are totally separate.**Common Issues:**

Why doesn't Database Heap Storage and Log Utilization Decline after a Mass Deletion?

The discussion here does not cover the database physical storage, which is subject to surprising behavior. When rows are added to the database, the heap storage files may need to expand to accommodate them, or the database may allocate additional files. After the rows are deleted, the data files do not necessarily decrease; rather, the space is marked as 'free' and will be reused when possible. It may be possible to reduce Database storage utilization, but it is an involved project which requires downtime. For institutions which are self-hosted, please consult a DBA. Managed Hosted institutions can contact Blackboard Client Support.