Introduction to Kaltura Batch Processes

Last Modified on 07/15/2021 4:11 pm IDT

The Kaltura batch management module implements a modular and distributed architecture, designed to answer the growing business and operational needs for site elasticity and smart distribution of system resources. The purpose of this document is to describe the architecture of the Kaltura batch management module with special emphasis on understanding the batch tasks and services that play a part in the Kaltura content ingestion flow.

What is Batch processing?

(From Wikipedia)
Batch processing is the execution of a series of programs ("jobs") on a computer without manual intervention. Batch jobs are set up so they can be run to completion without manual intervention, so all input data is preselected through scripts or command-line parameters. This is in contrast to "online" or interactive programswhich prompt the user for such input. A program takes a set of data files as input, process the data, and produces a set of output data files. This operating environment is termed as "batch processing" because the input data are collected into batches on files and are processed in batches by the program.

Kaltura Batch Task

A Kaltura Batch Task is a stand-alone task which is designed to be executed within the Kaltura Platform by a batch process. Kaltura batch tasks are initiated by a Kaltura API call that is triggered either by a specific end-user workflow or by an internal batch processing flow management entity.

When created, each batch task is stored within a dedicated data base record holding all information related to its specific type, its executing state, its priority and other operational information. For more information on batch tasks type classification, please refer to the Kaltura Batch Tasks Type Classification section.

Kaltura Batch Service

A Kaltura Batch Service is a configurable set of parameters defining a specific service that handles a batch task of a specific type in a specific way. A batch service is defined by parameters such as service name, the type of batch tasks it should handle, the name of the process that should be executed to operate the service, the maximum number of instances each service can operate at a given time, the execution schedule of the service and other operable parameters. There are 3 main types of batch services:

- **Batch Execution Service**
  A batch service that executes a full operation on a specific type of batch tasks.
- **Batch Closure Service**
  A batch service that only handles the finalization of a previous operation on a specific type of batch tasks.
- **Batch Periodic Service**
  A batch service that is mainly used for system maintenance operations, and does not handle batch tasks.

For more information on the Kaltura batch services, please refer to the Kaltura Default Batch Services section.

Kaltura Batch Process

A Kaltura Batch Process is one instance of a specific Kaltura batch service, executing the specific actions and logic needed for handling a specific type of batch tasks. Upon execution, each batch process checks for the next relevant pending batch task to be handled and operates on it.

Kaltura Batch Jobs API

A set of specific APIs used for implementing the internal and external flows related to the Kaltura batch processing implementation.

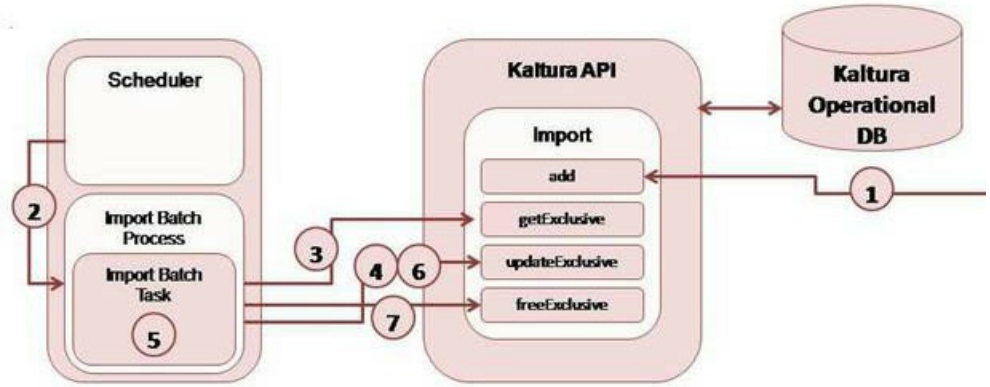
Kaltura Batch Scheduler

The Kaltura Batch Scheduler is a continual process, responsible for the scheduling of the batch services assigned to it. It schedules the execution of batch processes according to the load of pending batch tasks in the system and according to the scheduling rules defined in its configuration for the different batch services. The Kaltura batch scheduler is assisted by a special batch periodic service, named Scheduler Helper, providing the batch scheduler with relevant information on the current state of batch processes and batch tasks.

A Kaltura Batch Scheduler can run as a single scheduler within the platform deployment or run as one of many schedulers in a scaled-up platform configuration. The defined set of batch services controlled by each batch scheduler can be extended, reduced or adjusted in run-time according to system functional and scalability needs.

Internal Batch Processing of a Single Batch Task

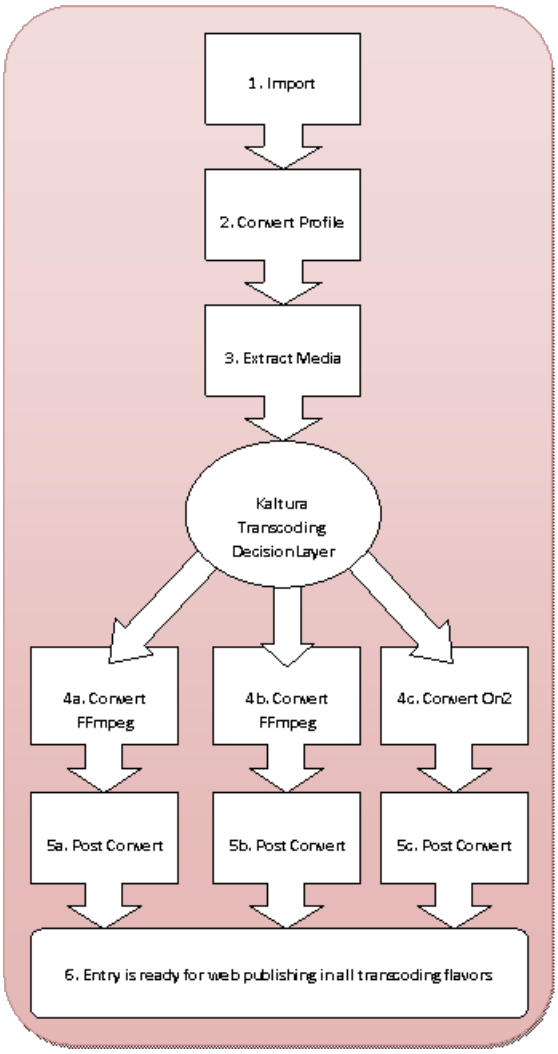
The following diagram illustrates the internal processing flow of a single batch task (import)



1. A new import task is added via Kaltura API as the first step of a content ingestion flow for a new rich-media file, following an end-user import action.
2. The Batch Scheduler executes a new batch process for executing the import job service.
3. The Import batch process asks for the next pending import task via Kaltura API.
4. The Import batch process updates the import batch task state to "Started".
5. The Import batch process transfers the rich-media file from its original location to the Kaltura platform.
6. The Import batch process updates the import batch task state to "Done".
7. The Import batch process releases the import batch task and ends.

Batch Processing Flow of a Successful Entry Ingestion

The following diagram describes the internal batch processing flow for full ingestion of rich-media files by the Kaltura online video platform - from import (detailed above) to full transcoding into various 'transcoding flavors' for playback. This is a simplified flow of a successful ingestion process.



1. The **Import batch process** transfers the new video file from its original location to the Kaltura platform
2. A **convert profile batch task** is created as a parent task to all the batch tasks related to the transcoding of the video file. An **extract media batch task** is created as well.
3. The **extract media batch process** extracts media related parameters from the headers of the video file that is about to be transcoded into web quality formats (flavors). This information is then passed over to the Kaltura **transcoding decision layer** for deciding on the optimal quality flavors and transcoding options to be used. Based on these decisions a suitable **convert batch task** is created for each one of the transcoding flavors to be generated.
4. Each **convert batch process** (4a, 4b, 4c) handles transcoding of the original media file into a specific transcoding flavor. In this example: 2 **convert batch tasks** are processed by**convert batch processes** that utilize the FFmpeg transcoding engine and one **convert batch task** is processed by a **convert batch process** that utilizes the On2 transcoding engine. Upon success, **post convert batch tasks** are

created

5. Each **post-convert batch process** (5a, 5b, 5c) processes the relevant **post convert batch task** for creating a thumbnail image and for extracting and storing media info about the created flavor for later use.
6. When all previously described **post convert batch tasks** have completed successfully, the new entry is available for web publishing in all of the required web quality flavors.

Kaltura Batch Tasks Type Classification

The following table lists the different types of batch tasks currently handled by the batch processing module.

Batch Task Type Classification (Internal Type ID)	Batch Sub Types Classification (Internal Sub Type ID)
Convert (0)	On2 (1)
	FFmpeg (2)
	Mencoder (3)
	Encoding.com (4)
	FFmpeg-Aux (5)
Import (1)	N/A
Flatten (3)	N/A
Bulk Upload (4)	N/A
Download (6)	N/A
Convert Profile (10)	N/A
Post Convert (11)	N/A
Extract Media (14)	Entry Input (0)
	Flavor Input (1)
Send Email (15)	Per email type
Send Notification (16)	Per server notification type

Kaltura Default Batch Services

The Kaltura online video platform includes a set of default batch services that are

required for system operation. The following table describes these services:

Service Name	Service System Name	Service Classification	Batch Tasks Handled By This Service	Description
Import Service	KAsyncImport	Batch Execution Service	Import	Handles the physical transferring of rich-media files imported by content managers and/or by end-users from their original location to the Kaltura platform
Bulk Upload Service	KAsyncBulkUpload	Batch Execution Service	Bulk Upload	Handles the processing of a bulk upload operation. Analyzes bulk upload csv and creates multiple import batch tasks to be processed separately
Bulk Upload Closer Service	KAsyncBulkUploadCloser	Batch Closure Service	Bulk Upload	Finalize bulk upload operation based on the completion status of all batch tasks related to the ingestion process of the uploaded files
Extract Media Service	KAsyncExtractMedia	Batch Execution Service	Extract Media	Extract media related information from media files to serve as an input for optimal transcoding operation
Convert Service	KAsyncConvert	Batch Execution Service	Convert	Handles the actual transcoding of one video file from one format to a specific quality flavor. Based on the transcoding requirements and system load, the convert service can operate transcoding action by utilizing one of the transcoding engines that are configured in the system.
Divert	KAsyncDivertConvert	Batch Execution	Convert	Handles real-time diversion of

Service Name	Service System Name	Service Classification	Batch Tasks Handled By This Service	Description
Conversion Service		Service		a convert task from one transcoding engine to another (specifically divert convert tasks to encoding.com if operable within the specific deployment and when needed for balancing system transcoding load)
Convert Closer Service	KAsyncConvertCloser	Batch Closure Service	Convert	Handles the finalization of a specific convert task (specifically handles the finalization of convert being handled by encoding .com or by a distributed scheduler)
Post Convert Service	KAsyncPostConvert	Batch Execution Service	Post Convert	Handles the last steps of a specific convert task including thumbnail creation and extracting media info from created flavors.
Convert Profile Closer Service	KAsyncConvertProfileCloser	Batch Closure Service	Convert Profile	Handles the finalization of in-progress convert tasks related to one entry when not all tasks were finalized before a defined timeout
Download Closer Service	KAsyncBulkDownloadCloser	Batch Closure Service	Download	Handles the completion of entry download flow, specifically responsible for triggering an email to the end-user with the download location
Mailer Service	KAsyncMailer	Batch Execution Service	Send Email	Handles all system generated emails sent by the Kaltura platform upon different events.
Notification Service	KAsyncNotifier	Batch Execution Service	Send Notification	Handles all server notifications sent by the Kaltura platform to web components (server/client) that are

Service Name	Service System Name	Service Classification	Batch Tasks Handled By This Service	Description
				integrated with The Kaltura notification system
Shared Imports Cleanup Service	DirectoryCleanupLocalImport	Periodic Batch Service	N/A	This is a scheduled maintenance service that cleans up the 'byproducts' of an import task
Shared Thumbnails Cleanup Service	DirectoryCleanupLocalThumb	Periodic Batch Service	N/A	This is a scheduled maintenance service that cleans up the 'byproducts' of a thumbnail creation process
Shared Converts Cleanup Service	DirectoryCleanupLocalConvert	Periodic Batch Service	N/A	This is a scheduled maintenance service that cleans up the 'byproducts' of a convert task
Database Cleanup Service	KAsyncDbCleanup	Periodic Batch Service	N/A	This is a scheduled maintenance that handles database cleanup
Scheduler Helper Service	KScheduleHelper	Periodic Batch Service	N/A	Handles all communication between the Batch Schedulers deployed in the Kaltura platform and the Kaltura API/DB

⚠ It is required that you batch the jobs for any bulk action. For any bulk actions that will create / edit / delete more than 5,000 entries or users, including Categories bulk uploads, please submit as batches of 500. If you are using the API, please batch as 500, sleep for 15 minutes, then submit the next batch of 500.