SM Sync Coordinator:
Super Admin Tool User Guide

Version: 1.2

Last Updated: November 11, 2019

Table of Contents

Introduction

SM Sync Coordinator is designed to import a variety of files, including JSON and CSV, into the SessionM Platform. This tool allows users to build templates that specify the processing instructions that are applied to a file or file set. Then, using a particular template, users can run an import for all uploaded files.

The heart of the SM Sync Coordinator ("SM Sync" in this document) is its templates, which are reusable configuration files that determine what types of files are to be imported and specify the rules that pertain to the import. Broadly speaking, a file can contain any type of ingestible data. But, the two primary types this tool ingests are CSV and JSON files. A file set can have one or more individual files in it, and each set is associated with an application by its API key.

Each file defined in the file set has a job type that specifies what type of import is required. For example, a file with a job type of "user" might contain a few basic details about the user such as their ID, email address and zip code. Alternatively, a file with a job type of "user_tag" might contain an external ID and a tag that assigns to the user a targeted customer segment.

A template also specifies a file's order in the processing queue for a file set. Each file is ingested using a specific filter that gets applied to each line of the file. Currently, the primary filter available is "copy", which instructs the importer to simply copy the file's content during processing.

In addition to its import functions, SM Sync provides several different pages that allow you to review the attributes of its templates, file sets and jobs.

SM Sync Workflow

The typical SM Sync workflow occurs in the following stages:

1. Create a template, or select an existing one.
2. Create and run a job, the core of which is to upload a file or files for ingestion.
3. Monitor a job's progress.
4. If you want, review details of the job and its associated file set.

Note that a job can comprise multiple smaller sub-jobs, called "importer jobs," that represent the chunks of data being processed. (The details for these importer jobs are also available for review.)

What's in this Guide

This guide provides the following sections:

Section	Description
Logging In	How to log in to SM Sync.
Set API key default	How to configure API key default
Getting to Know the Main Page	Summary information about SM Sync's main page.
Creating a Template	Procedural and conceptual instruction for creating a template.
Creating a Job	Procedural and conceptual instruction for creating a job.
Viewing Job and File Information	How and what to view for SM Sync jobs and files.
Managing Templates	Procedures for viewing, editing and deleting templates.

For more information on using SM Sync, see the *SM Sync Data Importers Guide*.

Logging In

Generally, in order to use SM Sync, the following information is provided:

- Super Admin role in the SMP
- API Key default configured

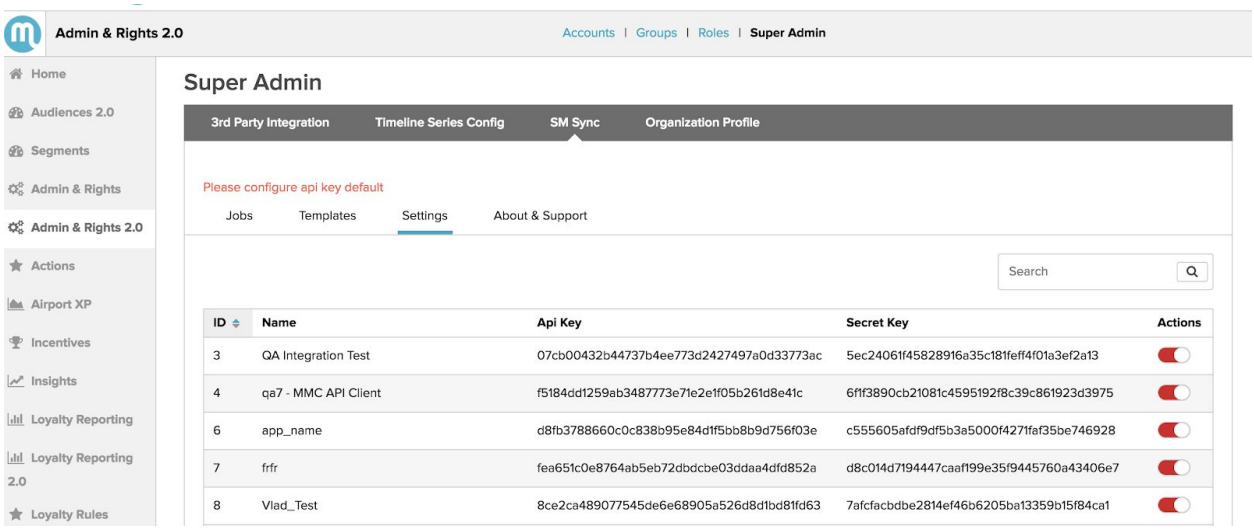
To login to SM Sync:

1. Access SessionM Platform (SMP) with Super Admin role.
2. Open Admin & Rights 2.0 module.
3. Click to the Super Admin Link and select SM Sync tool
4. SM Sync opens and displays the All SM Sync Jobs page, which is discussed in Getting to Know the Main Page. If the tool is used for the first time, the user will be redirected to the Settings tab where they should configure the default api key for the first time.

Set API Key Default

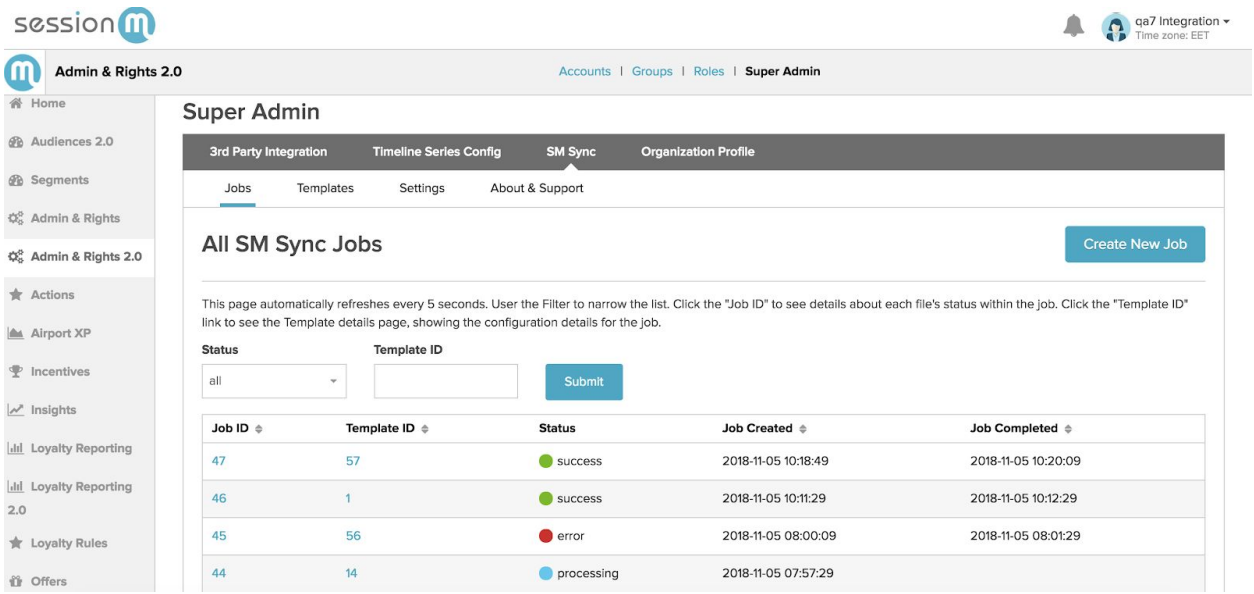
The first time a user wants to use SM Sync for an organization by accessing the link from Super Admin, they are redirected to the Settings tab. Here, they can select a default API key for the applications that will use the tool. The settings page displays all API keys for applications defined in the Digital Properties Module. The other tabs are disabled until this action is done (e.g., jobs or templates).

The application selected API key with toggle on is not accessible to be disabled. This is switched off when another application is configured as the default.



Getting to Know the Main Page

Upon logging in, the main page of SM Sync displays:



This page provides you with high level information about each job, such as its ID, the ID of its related template, its status and the date and times for the creation and completion of the job.

In the View Jobs table you can use the filter fields to search for and display specific SM Sync jobs. Specify values for any combination of the filter fields, which are detailed in the table below. Then click the Submit button to query for the jobs that meet your search criteria.

Filter Field	Description
Status	From the dropdown, choose one of the following values: ● *All* - Returns jobs that have run. ● *Error* - Returns jobs with a status of “error.” ● *Processing* - Returns jobs with the interim status of “processing.” ● *Success* - Returns jobs with a status of “success.”
Template ID	Enter the identifier for the template.

In addition to filtering for jobs, you can perform the following from the All SM Sync Jobs page:

- Create a new SM Sync job by clicking on the Create New Job button.
- View job and file information by clicking on a Job ID.
- View template information by clicking on a template ID.

Creating a Template

This section details the procedure for creating a template. The workflow is divided into two parts: specifying details for the template itself and defining how each file of a file set is processed.

One of the primary characteristics of a template is its strategy, which is how the files in the file set are processed for ingestion. The SM Sync importer provides three strategies:

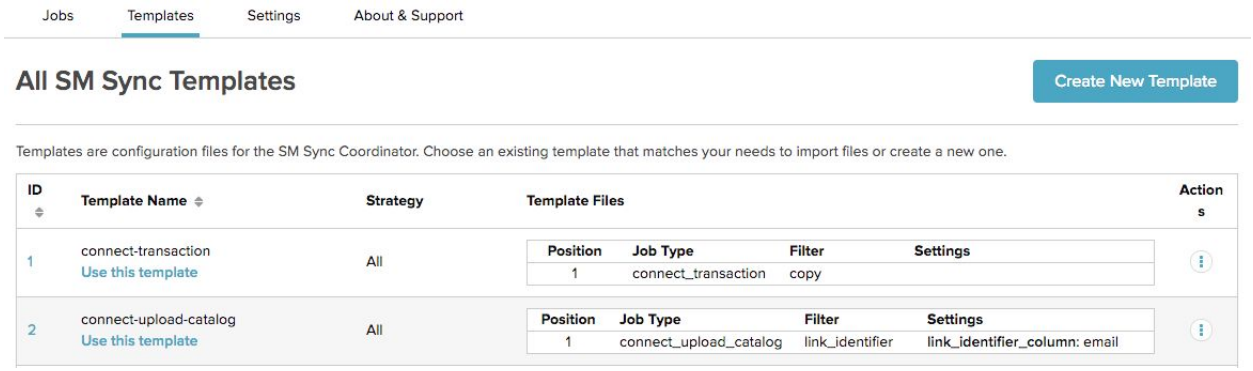
- “All” - Files are processed as they appear in the data store. This strategy ensures that files within the file set are processed sequentially. For example, consider a file set that contains two jobs, "user" and "user_tag.” In this case, the processing of the “user_tag” job cannot commence until the "user" job is finished. Note, however, that if you upload multiple file sets for this template, they can be processed concurrently.
- “Sequential” - Files are processed sequentially; each file handled only after processing for the previous one is complete. This strategy executes based on the file set’s timestamp, not the current time.
- “Stride” - Files are processed sequentially, but with two additional controls:
    - “Stride Duration” - Determines the duration (period of time) scheduled between imports, in hours, minutes or seconds.
    - “Offset Window” - Defines a percentage of the stride duration when files can be placed in the data store, or accepted, for processing. This percentage of the duration is an offset window that is in effect both before and after the scheduled import time. For example, consider your stride duration is set to 60m, and the offset window is set to 0.25. If the first file is uploaded at 12:00 (military time), then the second file must be imported between 12:45 and 13:15. In other words, because the stride is 60m, the offset of 0.25 equals 15 minutes, which is in effect both before and after the upcoming 1:00 import. More specifically, the offset window starts at 12:45 and ends at 13:15. Files can be imported anywhere in that duration.

  This strategy executes based on the file set’s timestamp, not the current time.

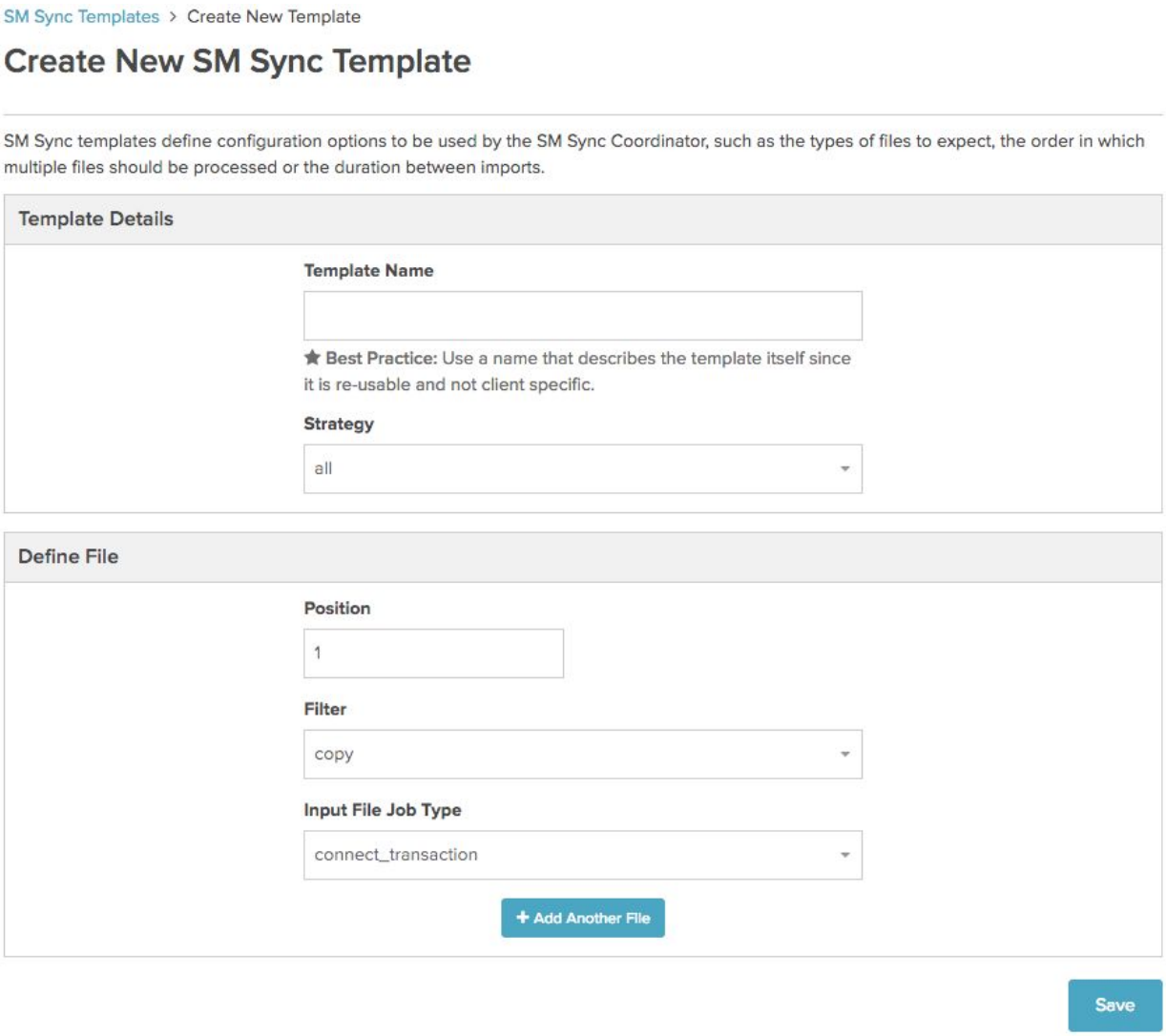
All files in the file set need to be present for the import process to start. Each file set can have only a single file of each type included in the set. In other words, it is invalid to upload a file set that contains two files of the same job type and file timestamp. For example, you cannot upload two files of the job type, “user.”

To create a template:

1. In the top menu, click the Templates option. The All SM Sync Templates page opens:



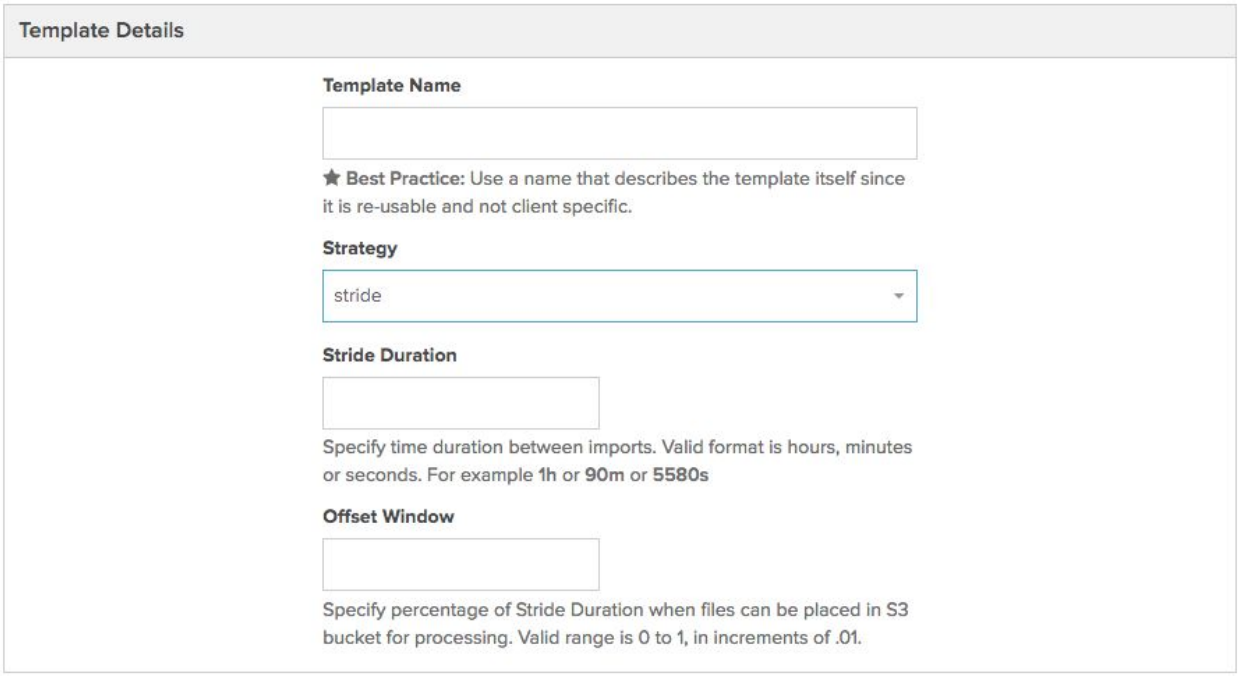
2. Click the Create New Template button. The Create New SM Sync Template page opens:



3. In the Template Details section of the page, enter a name for the template in the Template Name field. Names can contain only alphanumeric characters and dashes.

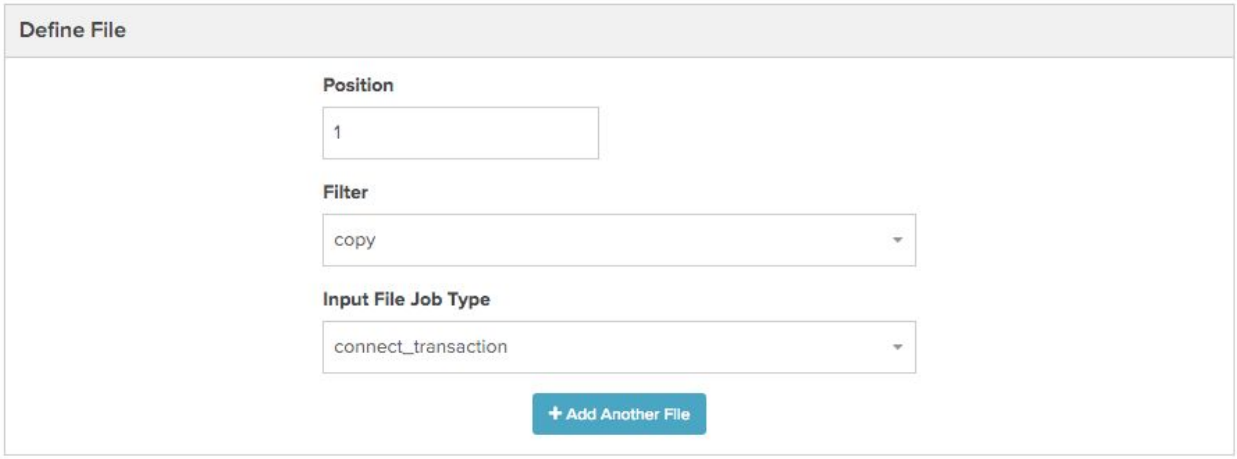
4. Then, from the Strategy dropdown, select the processing order to apply when multiple files sets are being uploaded: *All*, *Sequential* or *Stride*. For more information, see this section's introduction.

   If you chose *All* or *Sequential*, proceed to step 6; if you chose *Stride*, two additional fields are displayed in the Template Details section of the page:



5. In the Stride Duration field, specify duration between imports; in the Offset Window field, enter a value that defines a window for when the files can be placed in the data store for processing. For more information, see this section's introduction.

6. Scroll down to the Define File section of the page, which is shown below:



7. Use the Define File section of the page to define each file being processed in the file set.

    a. In the Position field, specify the file's place in the processing sequence.

b. In the Filter dropdown, select a preprocessing filter that can be applied to each line of the file. Currently, only two filters are supported: "copy," which copies the file from one bucket to another; and “link_identifier,” which copies a CSV column name specified with the placeholder value of “link_identifier.”
c. In the Input File Job Type, select the type of processing job required for the file being imported. For information, see the *SM Sync Data Importers Guide*.

8. If you need to define another file, click the Add Another File button and repeat step 7; otherwise, if you are done building the template, click the Submit button.

   You are taken to a page that displays the details for the new template and its associated file(s); the page also offers the Create New Job button to start an SM Sync job. Note that the new template is now listed on the SM Sync Templates page. Since the page lists templates in descending order by template ID, the new template is at the bottom.

With the template created, you can create and submit your SM Sync job.

Creating a Job

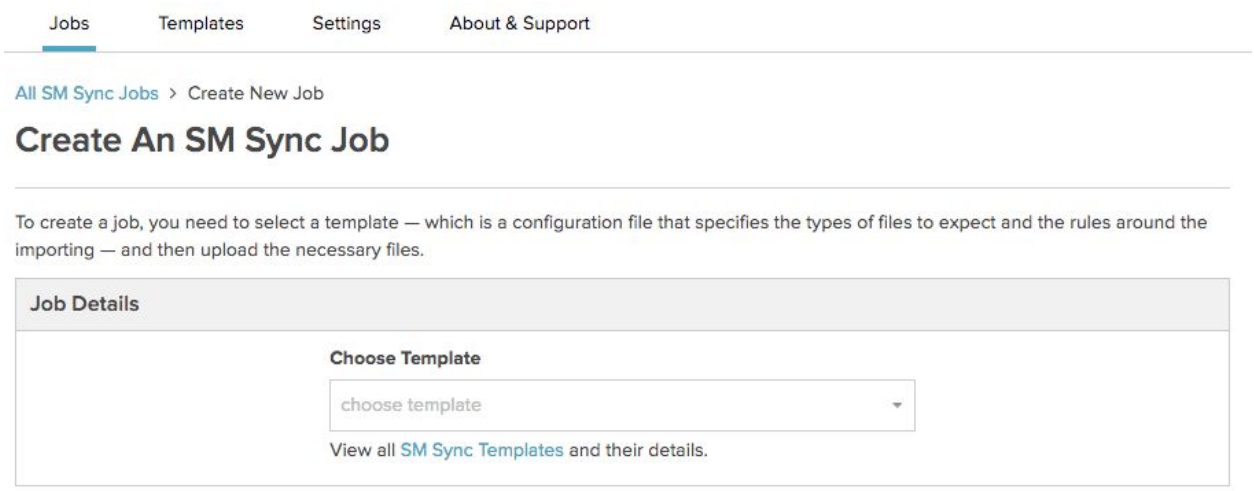
Assuming templates are available from the All SM Sync Templates page, you can create an SM Sync job. From the Create an SM Sync Job page, you can select the appropriate template, upload the corresponding file set and run the job.

To create a new job:

1. Access the Create an SM Sync Job in one of the following ways:

    a. From the top menu, click the Jobs option.
    b. From the All SM Sync Jobs page, click the Create New Job button.
    c. From the All SM Sync Templates page, click the “Use This Template” link for the template you want to use for creating an SM Sync job.

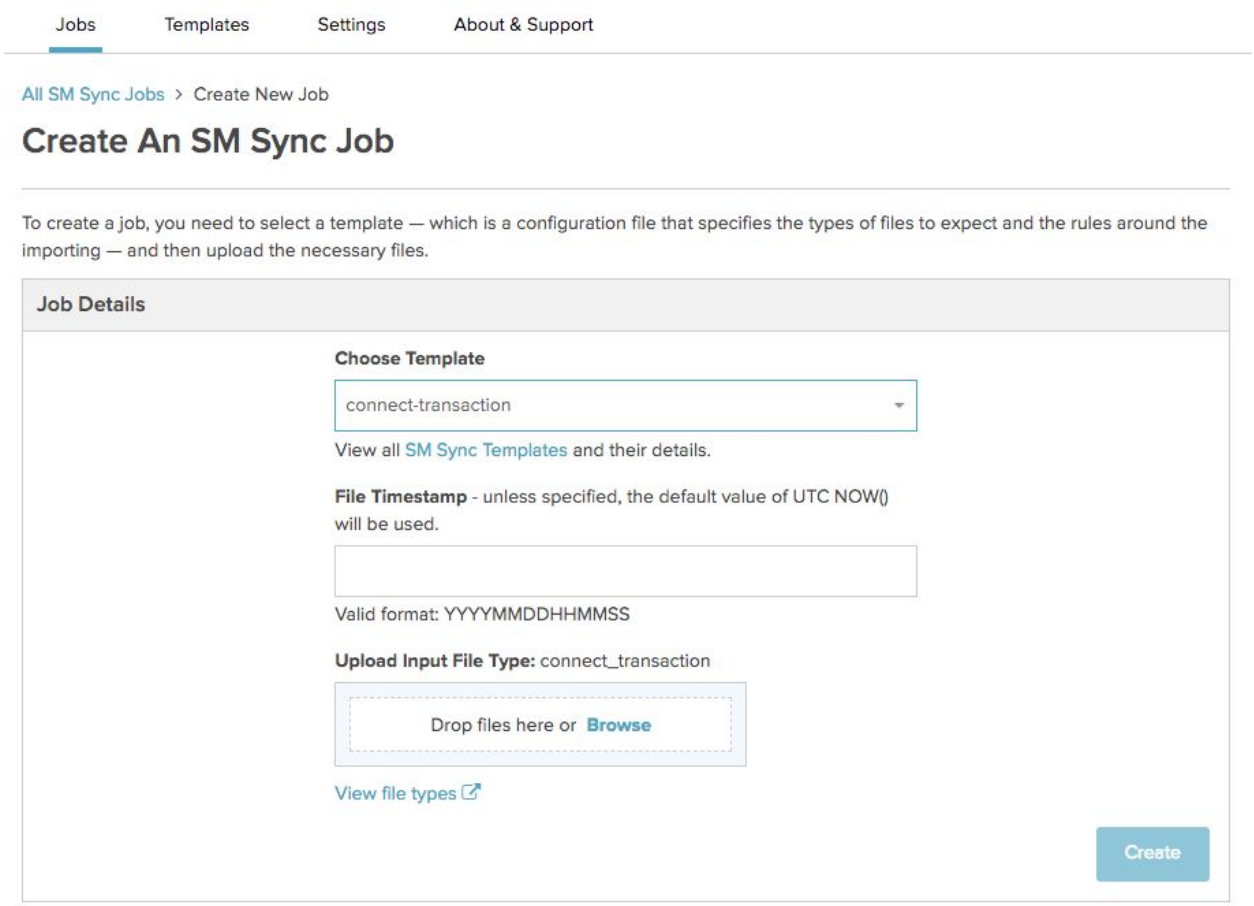
    The Create an SM Sync Job page opens:



2. In the Job Details section of the page, select a template from the Choose Template dropdown list.

    If you want to view specific details for the templates, click the SM Sync Templates link to review them on the All SM Sync Templates page.

    Once you’ve chosen a template, the Upload Input File Type and File Timestamp fields display:



Note that the Upload Input File Type field shows the event type associated with the template you selected.

3. Click the Browse button.

   The download dialog opens, allowing you to select the input file you want to upload.

4. Navigate to the file you want to upload and click the Open button.

5. Specify a value in the File Timestamp field; or, leave it blank and the system assigns a default value for the current timestamp.

6. Click the Create button.

   The new job displays in the View Jobs table on the All SM Sync Jobs page, showing a status of “processing” when the job is created. Then, once the job finishes, its status becomes “success” or “error” and a date and time value displays in the Job Completed column of the table. Note that runtimes for jobs can vary in duration, so be patient!

For more information on the All SM Sync Jobs page as well as its View Jobs table, see Getting to Know the Main Page.

Viewing Job and File Information

SM Sync provides a great deal of access to the details of each job and the files it ingests. The primary part of the UI that displays this information is the All SM Sync Jobs page. From it, you can review high level data about all the jobs or drill down into each job by clicking on its Job ID.

As you review this section, bear in mind the following relationships between jobs and files:

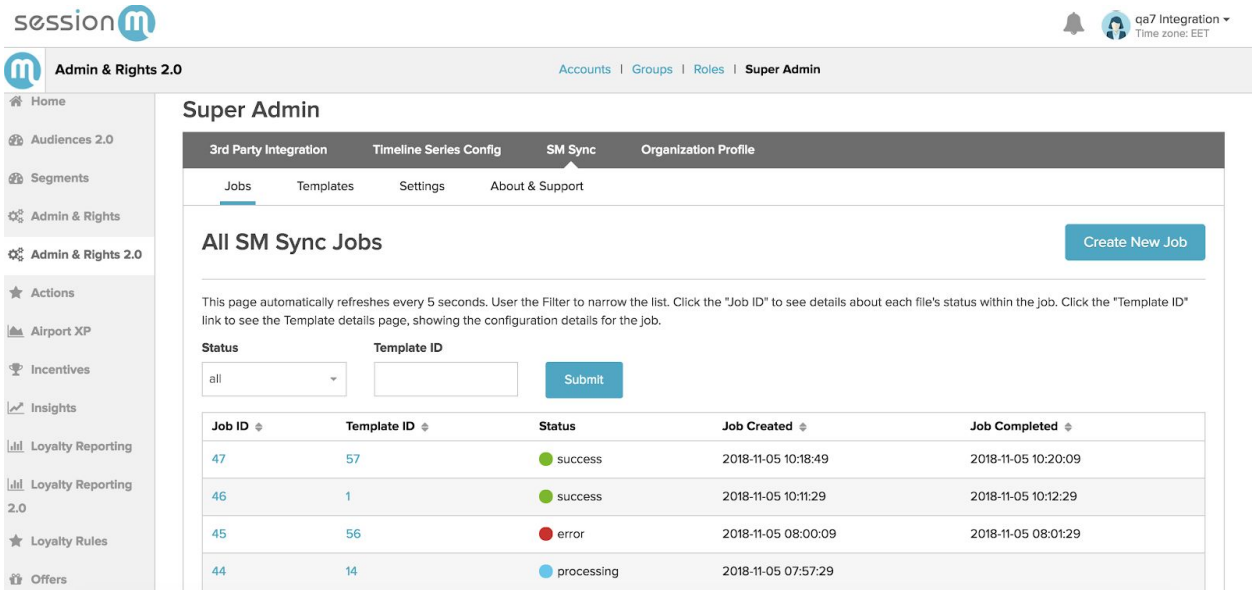
- A job contains one file set.
- Each file set contains one or more files.
- Each file can be processed for ingestion in one or more importer jobs.

Each importer job is a “chunk” of the file being ingested. As the importer runs, it divides the records in the file into discrete groupings. By default, the system splits the file into sets of 10,000 records. So if the file contained 10,002 records, SM Sync would create two importer job: one for 10,000 records and one for 2 records. For a CSV file, each of its lines is equivalent to one record. For a JSON file, the input must always be a JSON array; the number of records is equivalent to the number of elements in the array.

To view details on a specific job and the file(s) being ingested:

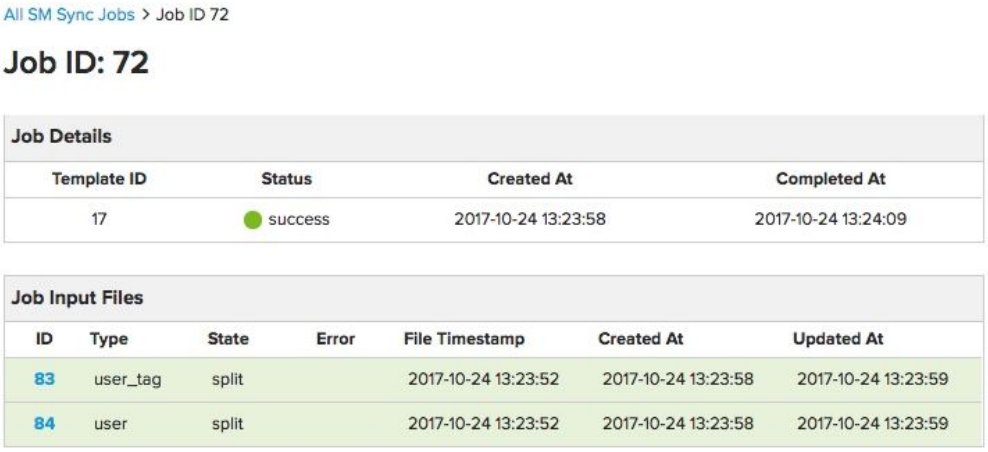
1. From the top menu, click the Jobs link.

   The All SM Sync Jobs page opens:



2. In the Job ID column of the View Jobs table, click the ID of the job you want to view.

   The Job ID page opens, showing the details and input files associated with the job:



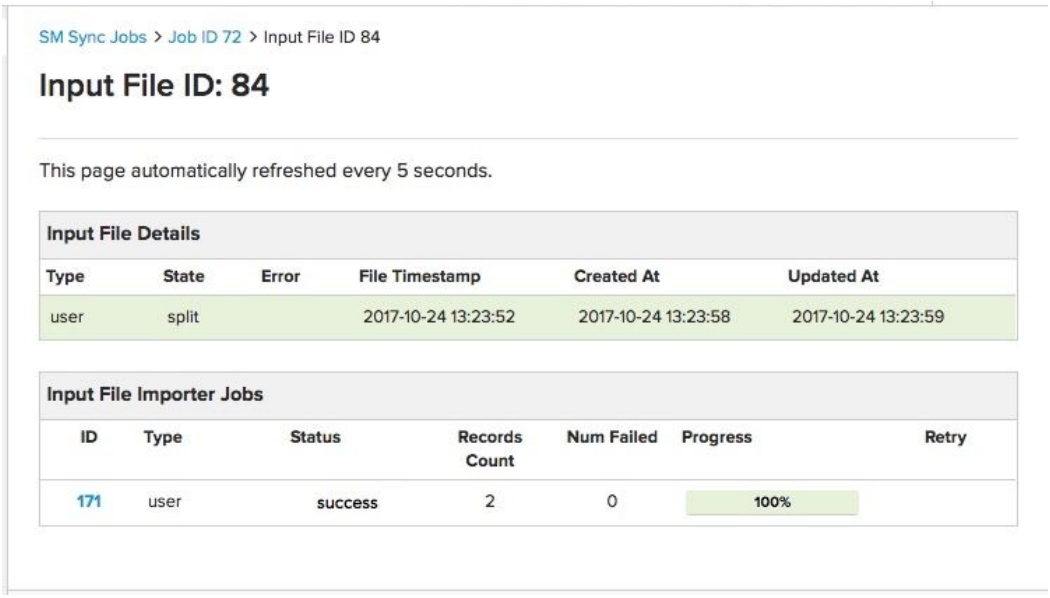
   In the Job Details table, each row displays the basic attributes of a job including its ID, status of the job ("processing," "success" or "error"), date and time job was created and date and time the job completed.

   In the Job Input Files table, each row displays the basic attributes of each file being ingested via the job. In the image above, two files have been ingested: "83," which is file containing user tag data; and, "84," contains user data. Attributes presented in this table include:

   a. *ID*, which is the identifier for the uploaded file.
   b. *Type*, which indicates the processing performed on the file. Possible values are documented in the *SM Sync Data Importers Guide*.
   c. *State*, which reflect the processing stage for the file set. Values include: "seen" for the initial state; "transferred" when the file begins being processed or split; and "split" for the final, processed state.
   d. *Error*, which is the error message associated with a failed job. Errors can occur for a variety of reasons. For example, improperly formed input file (e.g., JSON or CSV) or an encryption that used the wrong key.
   e. *File Timestamp*, which the file must contain to couple it with other members of the file set. Effectively, groups files together. Note that the file timestamp must be the same across the whole file set.
   f. *Created At*, which is the date and time the job was created.
   g. *Updated At*, which is the date and time the file was updated.

3. If you want more information on a specific job input file, click its ID in the ID column of the Job Input Files table.

   The Input File ID page opens, showing details about the ingested file and the smaller, "chunked" importer jobs that did the processing:



   In the Input File Details table, each row displays the basic attributes of a file in the file set. All of these attributes are discussed in step 2.

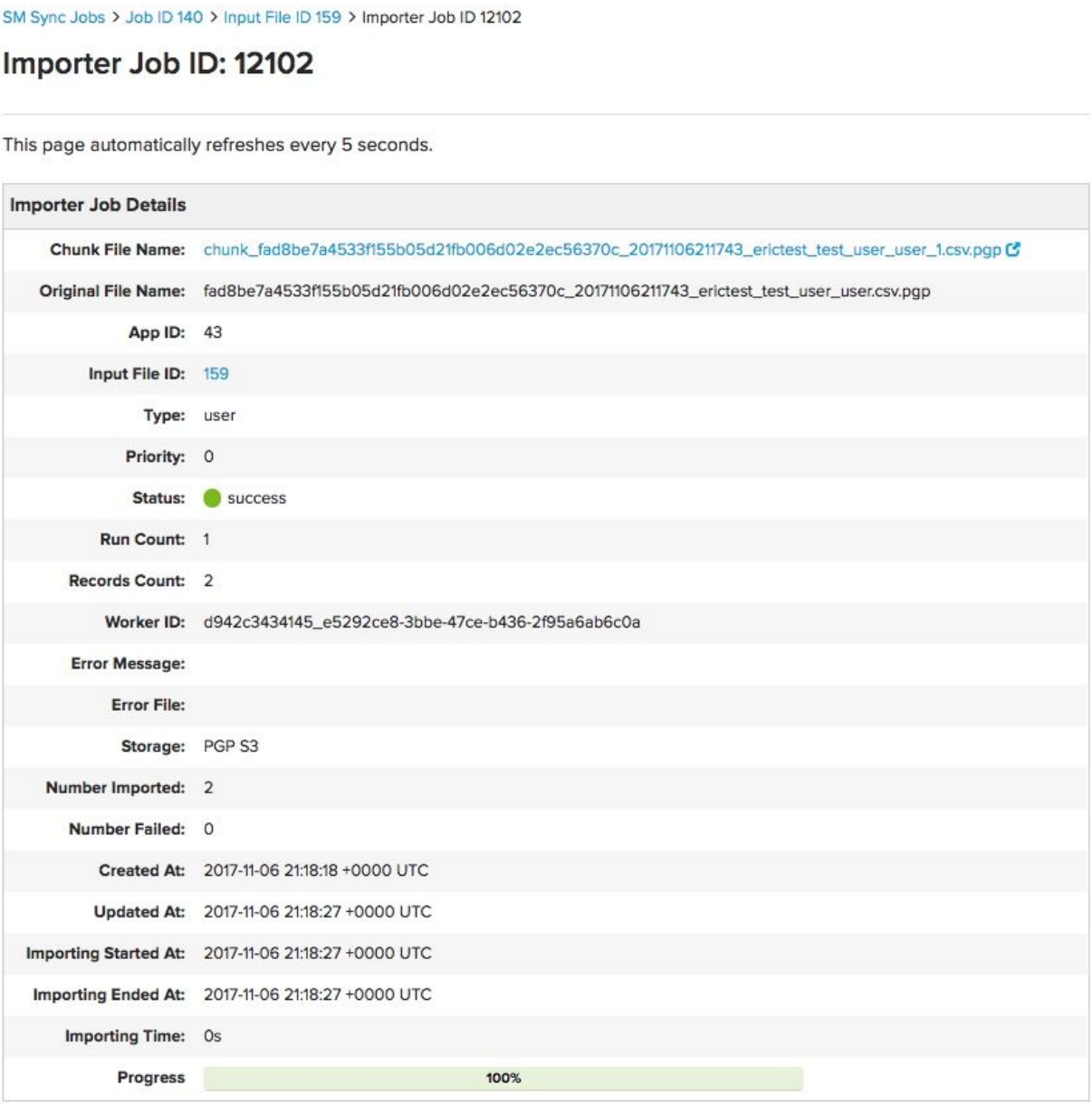
   In the Input File Importer Jobs table, each row displays the basic attributes of each smaller job that ran in order to process the file processed. These "importer jobs" represent the processed chunks of the data in the file with the following characteristics:

   a. *ID*, which is the identifier for the importer job.
   b. *Type*, which is the kind of processing performed on the file. Possible values are documented in the Types table of the *SM Sync Data Importers Guide*.
   c. *Status*, which displays "pending" when the job is created, "queued" when the job is scheduled by the import service, and "processing" when processing is underway. Then, once the importer job finishes, it displays the status of either "success" or "error."
   d. *Records Count*, which is the number of records in the chunk.
   e. *Num Failed*, which is the number of records in the chunk that were invalid or couldn't be processed.
   f. *Progress*, which is the percentage of one chunk that's been processed.
   g. *Retry*, which displays a Retry button if the chunk or the importer job fails. Click it to restart the job.

If you need more information on how SM Sync determines the appropriate number of importer jobs, see the introduction for this section.

4. If you want more information on an importer job associated with an input file, click the ID of the importer job in the ID column of the Input File Importer Jobs table.

   The Importer Job ID page opens:



Each row of the Importer Job ID table displays a different characteristic of importer job, including:

a. *Chunk File Name*, which is a link to the data in the chunk file.
b. *Original File Name*, which is the name of the original file before it's broken out into processing chunks. For information on the syntax for the name, see the "Input File Configuration" section in the *SM Sync Data Importers Guide*.
c. *App ID*, which is the application identifier.
d. *Input File ID*, which is a link to the Input File ID details page. For more information, see step 3 of this procedure.
e. *Type*, which is the type of processing defined for the job. For more information, see the *SM Sync Data Importers Guide*.
f. *Priority*, which defines importer job priority across the entire import service.
g. *Status*, which is the status of the importer job. It is "pending" when the job is created, "queued" when the job is scheduled by the import service, and "processing" when processing is underway. Then, once the importer job finishes, it displays the status of either "success" or "error."
h. *Run Count*, which displays how the number of times importer job has run.
i. *Records Count*, which displays the number of records in one chunk.
j. *Worker ID*, which is the unique identifier for the import service that owns the job.
k. *Error Message*, which displays an error message when importer job fails and has a status of "error."
l. *Error File*, which can contain a link to the file that contains all of the invalid records as well as why they were invalid. This error doesn't mean the importer job failed, only specific records inside the chunk file.
m. *Storage*, which storage was used. Currently can be "S3" and "FileSystem." In addition, the system supports PGP encrypted storage, which works in tandem with other storage. Value combine both types of storage; for example, "PGP S3." Note that if a PGP value is written here, all files have been encrypted.
n. *Number Imported*, which is the number of successfully imported records.
o. *Number Failed*, which is the number of records that failed to be imported.
p. *Created At*, which is the datetime the importer job was created.
q. *Updated At*, which is the datetime the importer job was updated.
r. *Importing Started At*, which is the datetime the importer job started.
s. *Importing Ended At*, which is the datetime the importer job ended.
t. *Importing Time*, which is the elapsed time.
u. *Progress*, which is the percentage of one chunk that's been processed.

5. If you want to view the actual file chunk, click the link that displays in the Chunk File Name row of the table.

   For example, if a CSV file is associated with a job type of “user,” the data chunk being imported might contain a user’s external ID and their email, as follows:

   ```
   external_id,email
   54855488,jsmith@sessionm.com
   59488986,jjones@gmail.com
   ```

   Or a CSV file associated with a job type of “event” might look like this:

   ```
   external_id,event_name,occurred_at
   sdangol771,event,2006-01-02T15:04:05-07:00
   ssmithl761,event,2006-01-02T15:04:05-07:01
   ```

Managing Templates

Templates are critical to setting up your workflow with SM Sync. As such, the UI provides the ability to view and manage existing ones.

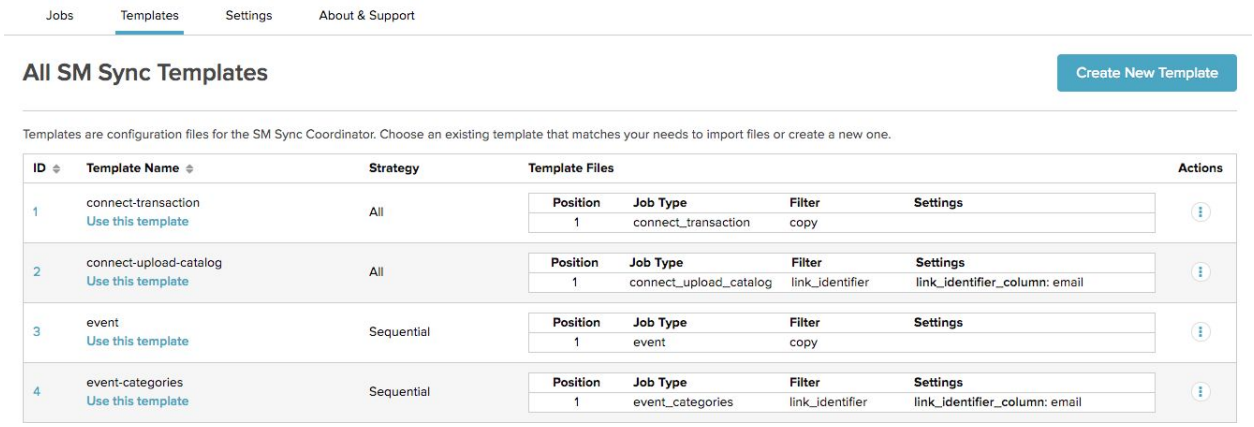
This section includes the following discussions:

- Accessing Templates
- Editing a Template
- Deleting a Template
- Viewing a Template in Detail

Accessing Templates

When you want to edit, delete or simply review summary information on SM Sync templates, access the All SM Sync Templates page.

To access all SM Sync templates, click the Templates link from the top menu. The All SM Sync Templates page opens:



In the table above, each row displays the basic attributes of each active template. Attributes include:

- *ID*, which is the identifier for the template.
- *Template Name*, which is the template's name. Next to the name is a Use This Template link. Clicking it brings you to the Create an SM Sync Job page, where you can create a job that is pre-populated with the template type of the named template. For more information, see Creating a Job.

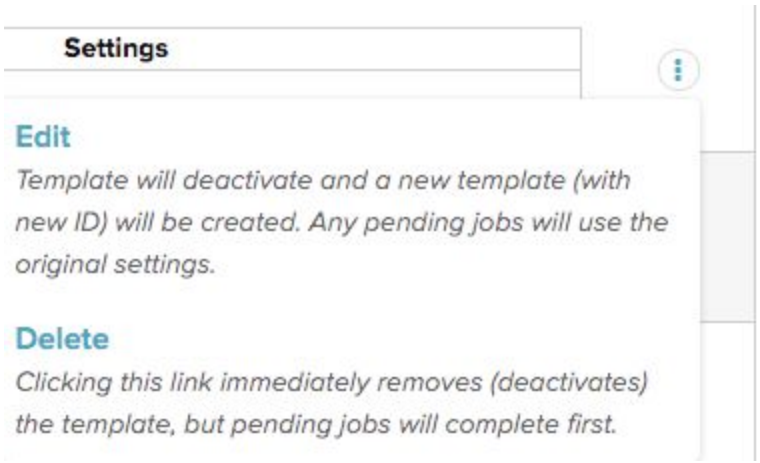
- Strategy, which is how the files in the file set are processed for ingestion. Values include: *All*, *Sequential* and *Stride*. For more information, see the discussion in Creating a Template.
- *Template Files* attributes, which include the following:
    - *Position* - Processing position relative to any other files.
    - *Job type* - Type of processing defined for the template. For more information, see the *SM Sync Data Importers Guide*.
    - *Filter* - Currently provides only two filters: "copy," which copies the file from one bucket to another; and “link_identifier,” which copies a CSV column name specified with the placeholder value of “link_identifier.”
    - *Settings* - Displays settings defined for a template file.
- *Actions*, which provides the ability to edit or delete the related template. For more information, see Editing a Template or Deleting a Template.

Editing a Template

As you work with SM Sync, you may find it necessary to modify some of a template’s existing details or attributes. Perhaps you need to alter the template’s strategy or its position in the processing queue. Note that when you do edit an existing template, the old version of the template is deactivated and a new one (with a new ID) is created in its place. Any jobs pending when you edit the template are processed using the original template details.

To edit a template:

1. From the All SM Sync Templates page, click the Actions symbol (three vertical dots) for the template you want to edit. The following menu appears:



2. Click the Edit link.

The template displays in the Create New SM Sync Template, populated with its existing values.

3. Modify the template or add another file to the template using the Add Another File button. Then click the Submit button.

   You are taken to a page that displays the details for the revised template and its associated file(s); the page also offers the Create New Job button to start an SM Sync job. Note that the revised template is now listed on the SM Sync Templates page. Since the page lists templates in descending order by template ID, the revised template, which has a new ID, is at the bottom.

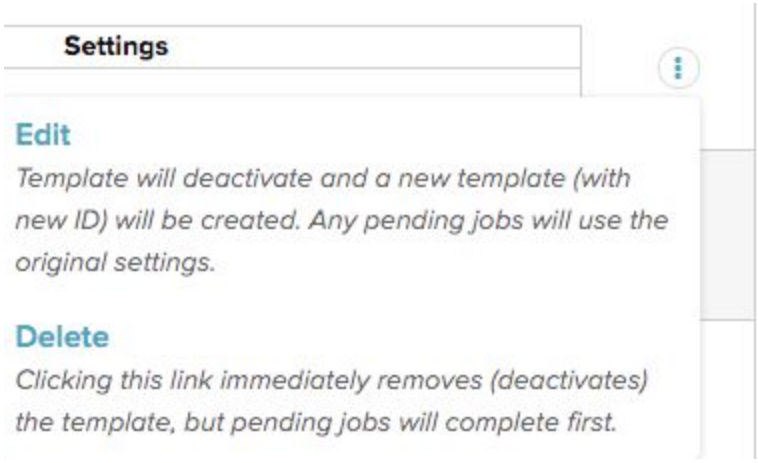
Deleting a Template

As you work with SM Sync, you may find it necessary to remove, or deactivate, an existing template so it doesn't get used in any future jobs. Any jobs pending when a template is deactivated are processed using the original template details.

To delete a template:

1. From the All SM Sync Templates page, click the Actions symbol (three vertical dots) for the template you want to delete:

   The following menu appears:



2. Click the Delete link.

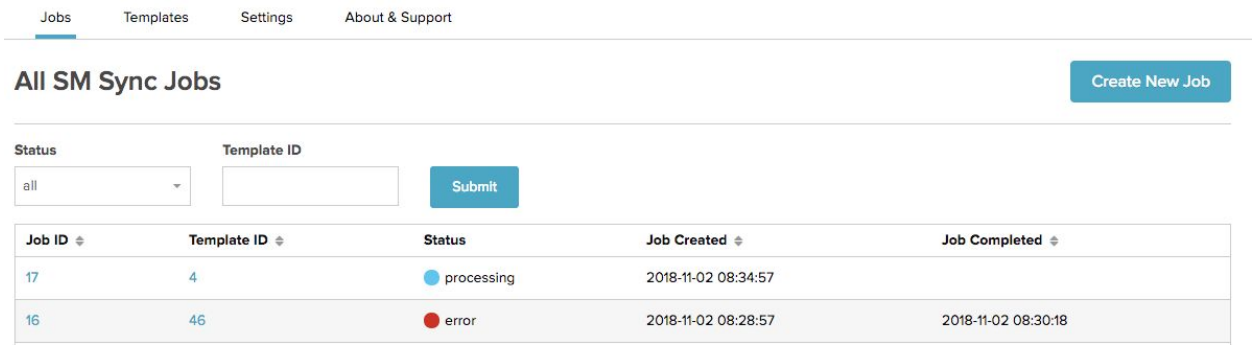
   The template is deactivated and removed from the All SM Sync Templates page.

Viewing a Template in Detail

SM Sync allows users to easily view details about a template and each of the files in its associated file set. Note that this section of the guide provides an extensive discussion on template attributes, but you can also review a summary treatment instead which is available in Accessing Templates.

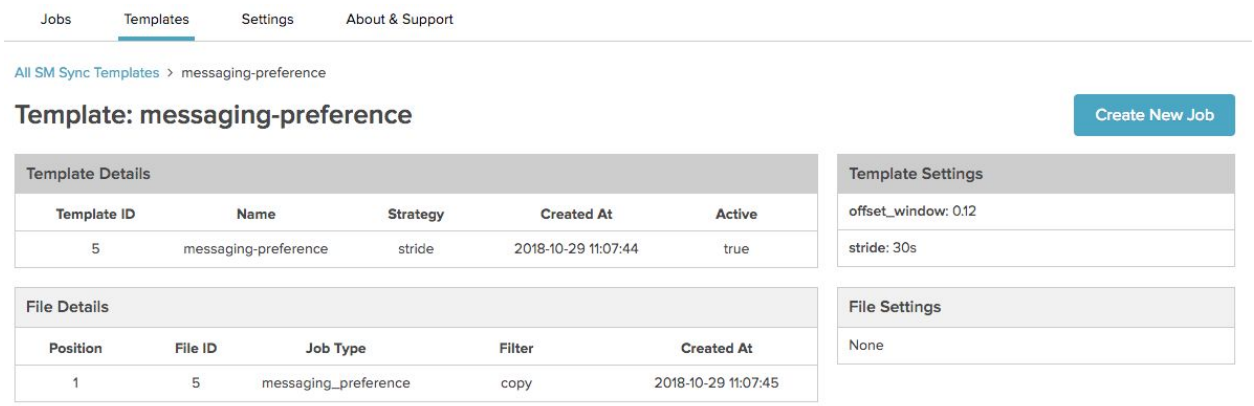
To view details on a specific template, click the Jobs link from the top menu.

The All SM Sync Jobs page opens:



Then, in the Template ID column of the View Jobs table, click the ID of the template you want to view.

The Template page opens:



The Template Details table displays the basic attributes of a template, including its ID, name, strategy, datetime at which it was created, and whether or not the template is active. Adjacent to this table, the Template Settings table indicates any specific settings associated with the

template. For example, since the strategy for the template is “stride,” the settings that display include “offset_window” and “stride.” For more information on strategy, see the Introduction in Creating a Template.

The bottom half of the page displays one or more File Details tables, each of which shows the basic attributes of the file being processed in the template, including:

a. *Position* - Processing position relative to any other files.
b. *File ID* - Identifier for the file.
c. *Job type* - Type of processing defined for the template. For more information, see the *SM Sync Data Importers Guide*.
d. *Filter* - Currently provides only two filters: "copy," which copies the file from one bucket to another; and “link_identifier,” which copies a CSV column name specified with the placeholder value of “link_identifier.”
e. *Created At* - Datetime at which template was created.

Adjacent to the File Details table, the File Setting table indicates any specific settings associated with the file. For example, settings for filters or jobs. Currently, the only supported settings are “deactivate_venues” (true or false) and “link_identifier_column,” which is a string.