

SingleStore XL Ingest

01/08/2025

SingleStore XL Ingest	3
Overview	3
Installation	3
Select Tables	3
Split Table into Slices	4
For Large Tables	4
For Partitioned Tables	5
For Smaller Tables	5
Configuration of XL Ingest	5
Dashboard	5
Configuration	6
Table Setup	7
Table Status	8
Table Slicing Status	8
Schedule Tasks	9
Task Status	10

SingleStore XL Ingest

Overview

SingleStore XL Ingest (“XL Ingest”), a component of SingleStore Flow, is companion software to SingleStore Ingest (“Ingest”). When large tables (greater than 10GB) are included in Ingest, a full extract may not be feasible due to long processing times. This increases the likelihood of encountering issues, and if a problem occurs, the entire initial extract must be rerun. Therefore, XL Ingest is essential for working with large tables.

XL Ingest handles the initial transfer of large tables by dividing them into smaller logical partitions. It then transfers multiple partitions from the source to the target in parallel. This ensures the transfer happens within a reasonable amount of time. Use XL Ingest to transfer the identified logical partitions of large tables in parallel to SingleStore for storage and processing.

To transfer large tables from source database to the desired destination database, perform the steps outlined in [Select Tables](#). Complete these steps before triggering an XL Ingest job to prevent data loss during the transition.

Note: Pausing updates on the source database is not required during this process. Both Ingest and XL Ingest can operate independently and concurrently without disrupting the source data.

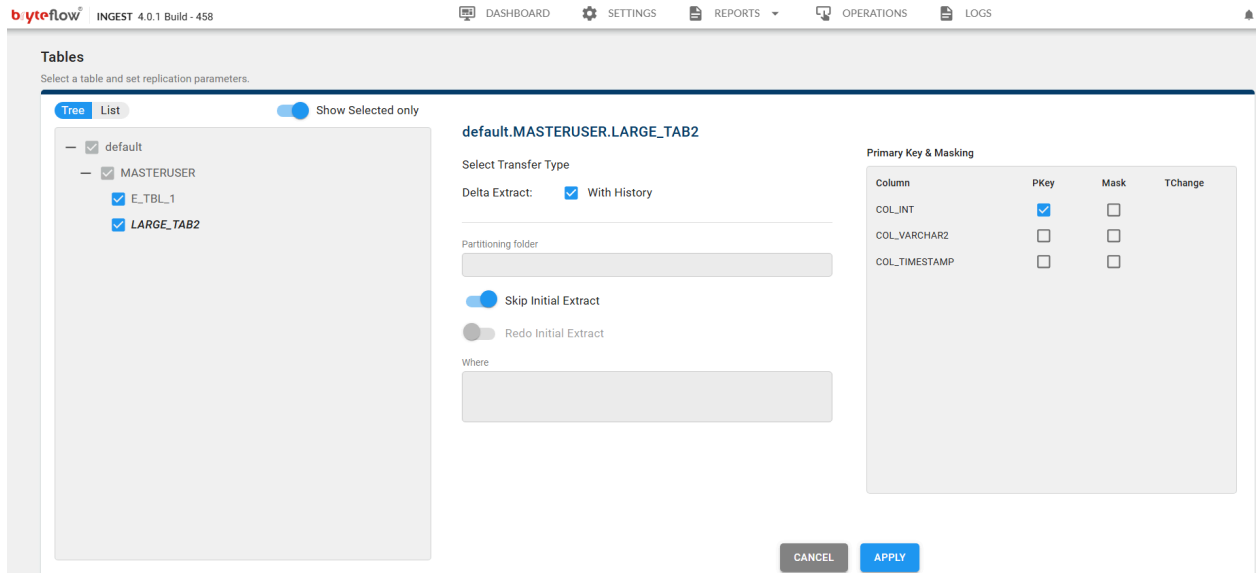
Installation

For details on how to install XL Ingest and other Flow components, refer to [SingleStore Flow Installation](#).

Select Tables

To ingest large data table(s) from source database to destination database, perform the following steps in *Ingest* before triggering a job in XL Ingest:

1. Navigate to **Dashboard > Tables**, and then select the gear icon.
2. Define a primary key (**Pkey**) and any necessary partitions for the table.



3. Enable **Skip Initial Extract** to bypass the initial extract and directly proceed with the delta load.
4. Select **Apply** to save the changes.
5. Navigate to **Dashboard > Operations** and disable the Ingest scheduler to ensure all tables are moved at the same time to the destination.
6. Initiate **Full Extract** to trigger the initial bulk load for all the selected tables, except those marked as **Skip Initial Extract**.
7. Initiate **Sync New Tables** to trigger the initial bulk load for tables marked as **Redo Initial Extract** and newly added tables in an ongoing replication. This captures the watermark for CDC and creates the table in the destination database.
8. Enable the Ingest scheduler in Ingest after transferring tables using XL Ingest.

Note: After marking tables with **Skip Initial Extract**, the next scheduled delta run automatically captures CDC for all tables, including those loaded with XL Ingest. XL Ingest prevents duplication during the CDC load by using a watermark to track changes.

Split Table into Slices

Large tables must be divided into notional slices based on the value of a single slice column, for example, Primary Key. For automatic slice determination, XL Ingest uses parameters like the number of slices needed and how many characters from the start of the slice column value must be used.

For Large Tables

XL Ingest automatically determines the slices based on the specified parameters. For example, a slice column like names can be divided by the first 3 characters. Alternatively, you can manually enter the slice values instead of using auto-slice.

For Partitioned Tables

Counting records is not necessary. XL Ingest automatically determines the list of partitions to create, with each partition treated as a slice.

For Smaller Tables

Slicing may not be necessary. The entire table can be processed as a single slice.

Configuration of XL Ingest

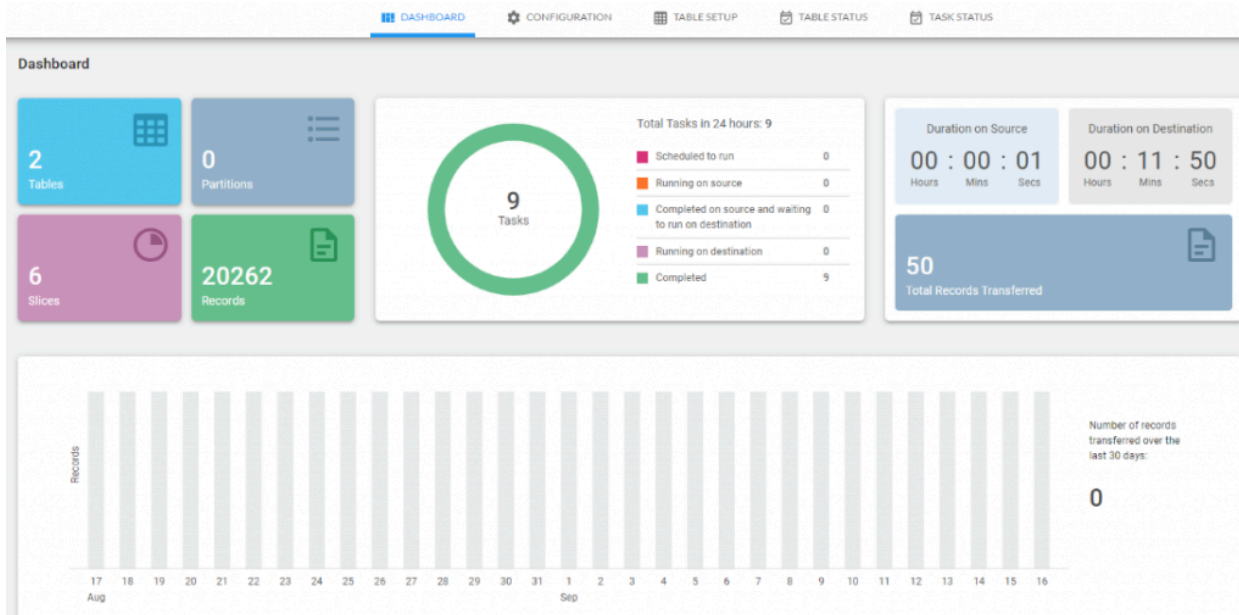
You can configure XL Ingest through the web console. To access the console, enter the following URL in Chrome: `localhost:8084`

The screen displays the following tabs:

- Dashboard
- Configuration
- Table Setup
- Table Status
- Task Status

Dashboard

Navigate to the **Dashboard** tab. The dashboard provides a summary of the data being managed and the work completed in the last 24 hours. It includes a pie chart showing task statuses and a bar chart indicating the amount of data transferred over the last 30 days.



Configuration

Navigate to the **Configuration** tab. Configure the source and destination database in Ingest. Go to <http://localhost:8084>. Although the setup is automatic, you must follow certain steps to transfer the large table to the desired destination. Set up XL Ingest by entering the following configuration details:

1. **Location of CDC Installation:** Enter the path for Ingest.
2. **Web Port:** Enter the port of XL Ingest (8084).
3. **No. of source threads:** Enter the number of source threads.
4. **No. of destination threads:** Enter the number of destination threads.
5. **Product Id:** By default, the product ID is the field.
6. **Licence Key:** Enter the license key.
7. **Expiry:** Displays the expiry date of your license.
8. **Delete on Sync:** Enable or disable as per the requirement.

Select **Save** to save the settings.

Configuration
Set up your XL Ingest configuration.

Location of CDC installation*: /ingest

Web Port: 8084

No. of source threads: 2

No. of destination threads: 2

Product Id*: X23-ANWNO-TFEDA-LPORI-F6WAP-PHPQM-9HADC

Licence Key*: B18HQ1RUF8P0EBHTZAZJMLVYA-FDFQA

Expiry*: Expires on 01 Jan 2025

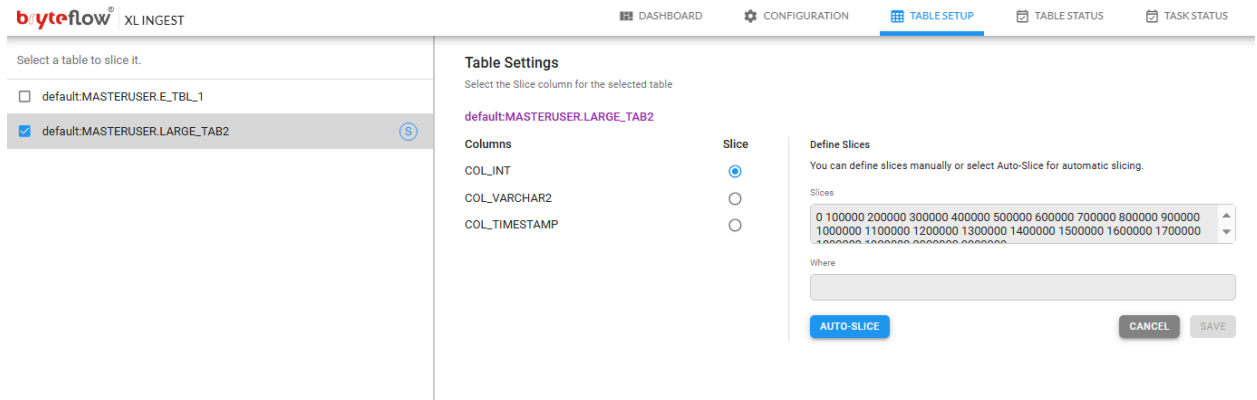
Delete on Sync

CANCEL SAVE

Table Setup

Table setup allows you to configure the tables.

1. Navigate to the **Table Setup** tab.



2. Select the tables from the list on the left to be processed in XL Ingest.
3. In the **Table Settings** page, choose the slicing column (usually the Primary Key). For partitioned tables, slicing is automatically handled by XL Ingest.
4. Select the **Auto-slice** button to automatically generate slices for large tables. Select the **Schedule Count** button to get the number of records in the table, enter the following auto-slice parameters:
 - a. **No. of Slices**: Number of slices to split the table into.
 - b. **Round-off**: Number of trailing significant digits to ignore when estimating the slice size. Simplifies the calculation by focusing only on the most significant digits of the total record count.

Auto Slicing

×

Slice Column: COL_INT

Data Type: NUMBER

Click on the Schedule Count button to get the table statistics.

Count	Max	Min
2M Recs	2000000	1
Size	Time	
118MB	02m48s	

SCHEDULE COUNT

Click on the Schedule Slice button after the number of records appear and slicing parameters are defined.

No. of Slices	Records per slice	Slice size
<input type="text" value="20"/>	100K Recs	5.9MB
Round-off (digits)	Time per slice	
<input type="text" value="4"/>	08s	

SCHEDULE SLICE

CANCEL

5. The **Schedule Slice** button is enabled only after the count is determined.

Table Status

After selecting and configuring the tables, the tables appear in the **Table Status** page.

1. Navigate to the **Table Status** tab.
2. It displays the status of each table and a bar graph indicating the amount of data transferred.
3. Select a table to view its slicing status and progress.

Table Slicing Status

Table Slicing Status page displays the list of all slices or partitions of the table.

Table Slicing Status

View the status of the slices within the table.

COMPARE FORCE COMPARE SYNC FORCE SYNC

Table Name	Source Records	Dest. Records	Variation	Slices	Status	Progress	Last Checked
default:MASTERUSER.LARGE_TAB2	99,999	99,999	0%	21	Unknown	<div style="width: 5%;"></div> 5%	

Compare & Sync

Under the Compare column you can compare the data in the source with that in the destination. Under the Sync column you can sync and update data in the destination with that of the source. You have the option to cancel a Compare or Sync action if it has not yet started.

Slice Name	Src. Records	Dest. Records	Status	COMPARE SELECTED	SYNC SELECTED
0 to 100000	99,999	99,999	✔	<input type="checkbox"/> SCHEDULE COMPARE	<input type="checkbox"/> SCHEDULE SYNC
100000 to 200000	-	-	🔄	<input type="checkbox"/> COMPARE NEEDED	<input type="checkbox"/> CANCEL SYNC
200000 to 300000	-	-	▲	<input type="checkbox"/> COMPARE NEEDED	<input type="checkbox"/> SCHEDULE SYNC
300000 to 400000	-	-	▲	<input type="checkbox"/> COMPARE NEEDED	<input type="checkbox"/> SCHEDULE SYNC
400000 to 500000	-	-	▲	<input type="checkbox"/> COMPARE NEEDED	<input type="checkbox"/> SCHEDULE SYNC
500000 to 600000	-	-	▲	<input type="checkbox"/> COMPARE NEEDED	<input type="checkbox"/> SCHEDULE SYNC
600000 to 700000	-	-	▲	<input type="checkbox"/> COMPARE NEEDED	<input type="checkbox"/> SCHEDULE SYNC
700000 to 800000	-	-	▲	<input type="checkbox"/> COMPARE NEEDED	<input type="checkbox"/> SCHEDULE SYNC
800000 to 900000	-	-	▲	<input type="checkbox"/> COMPARE NEEDED	<input type="checkbox"/> SCHEDULE SYNC

- Following are the buttons for each slice to schedule:
 - Compare:** Schedules a comparison of the slice between the source and destination database.
 - Force Compare:** Forces a comparison of the slice between the source and destination database, if it has been compared earlier.
 - Sync.** Schedules the transfer of the data between source and destination database.
 - Force Sync:** Forces the transfer of the data between the source and destination database, if the data has already been synced previously.
- The color of each button indicates the state of the respective task as follows:
 - Dark Blue:** Task can be scheduled and is recommended.
 - Light Blue:** Task can be scheduled but no urgent action is required.
 - Red:** Task is scheduled and can be cancelled if needed before it runs.
 - Grey:** Task is in progress.

Schedule Tasks

XL Ingest tasks are long-running and cannot be executed immediately. They must be scheduled, and the progress is available after execution. Perform the following tasks:

- Select **Schedule Compare** for the slice to start the slice comparison between the source and destination database.
- Select **Schedule Sync** for the slice to start the synchronization of data between the source and destination database.
- View the status of scheduled tasks in the **Task Status** page.

Task Status

Task Status page displays the list of the tasks in the last 24 hours.

1. Navigate to the **Task Status** tab.

The screenshot shows the 'Task Status' page in the Byteflow interface. The page title is 'Task Status' with a subtitle 'Status of actions taken in the last 24 hours'. The navigation bar includes 'DASHBOARD', 'CONFIGURATION', 'TABLE SETUP', 'TABLE STATUS', and 'TASK STATUS'. The filter bar shows 'Completed', 'Scheduled', 'Running', and 'Failed' all selected. The table below lists 12 tasks with the following data:

Table	Action	Slice	Records	Date	Status	Duration	Cancel Action*
default:MASTERUSER_LARGE_TAB2	Sync	7 of 21	100,000	2024-12-02 09:45:36	Scheduled	11s	CANCEL
default:MASTERUSER_LARGE_TAB2	Sync	6 of 21	100,000	2024-12-02 09:45:36	Running	02s	CANCEL
default:MASTERUSER_LARGE_TAB2	QComp	5 of 21	100,000	2024-12-02 09:45:40	Completed	02s	CANCEL
default:MASTERUSER_LARGE_TAB2	Sync	5 of 21	100,000	2024-12-02 09:44:54	Completed	09s	CANCEL
default:MASTERUSER_LARGE_TAB2	Sync	4 of 21	100,000	2024-12-02 09:44:50	Completed	13s	CANCEL
default:MASTERUSER_LARGE_TAB2	Sync	3 of 21	100,000	2024-12-02 09:44:23	Completed	14s	CANCEL
default:MASTERUSER_LARGE_TAB2	Sync	2 of 21	100,000	2024-12-02 09:41:29	Completed	09s	CANCEL
default:MASTERUSER_LARGE_TAB2	Sync	1 of 21	99,999	2024-12-02 09:37:31	Completed	11s	CANCEL
default:MASTERUSER_LARGE_TAB2	Sync	4 of 21	100,000	2024-12-02 09:34:17	Failed	02s	CANCEL
default:MASTERUSER_LARGE_TAB2	Sync	3 of 21	100,000	2024-12-02 09:34:05	Failed	02s	CANCEL

*A task can only be cancelled if it is yet to start.

2. Tasks are categorized as follows:
 - a. Completed
 - b. Scheduled
 - c. Running
 - d. Failed
3. Tasks can be filtered and monitored for progress. You can cancel scheduled tasks before they start, if necessary.