SingleStore Flow Installation

01/09/2025

SingleStore Flow Installation

Overview

The SingleStore Flow (“Flow”) system allows you to migrate data and do continuous change data capture (CDC). Flow has two primary components:

- **Ingest**: Ingest transfers the schema for all tables, moves table data for small tables (typically up to 10GB in size), and handles ongoing CDC of changes for all tables into SingleStore.
- **XL Ingest**: XL Ingest handles the initial transfer of large tables by breaking these tables into smaller logical partitions. It then copies multiple partitions from the source to the target in parallel. This helps transfer large tables in a reasonable amount of time.

To use Ingest and XL Ingest, install the products and, once installed, you may either purchase them or use the Developer trial license.

Prerequisites

System Requirements

- **VM requirements**: Minimum of 2 vCPUs, 8 GB RAM, and 100 GB of free disk space (or twice the size of the largest table you plan to extract with Ingest or XL Ingest). The VM must be close to the source database system to minimize latency.

Browser Requirements

- Use Chromium based browsers (Google Chrome or Microsoft Edge).

Java Installation

- Download and install Amazon Correto Java 21.

Prerequisites for Supported Source Databases

Oracle

Flow supports on-premises Oracle database and Amazon Oracle RDS database as source database.

On-Premises Oracle Database

Supported Oracle Versions

- Oracle 9i and above.

Enable Change Tracking on Oracle Database

To enable change tracking on an on-premises Oracle database, ensure the following configurations:

1. Ensure the Oracle database is running in `ARCHIVELOG` mode.
2. Supplemental logging must be enabled at the database level to ensure additional details are logged in the archive logs. Run the following command to enable full supplemental logging:

   ```
   ALTER DATABASE ADD SUPPLEMENTAL LOG DATA (ALL) COLUMNS;
   ```

   Alternatively, to enable minimal supplemental logging, run the following commands:

   ```
   ALTER DATABASE ADD SUPPLEMENTAL LOG DATA;
   ALTER DATABASE FORCE LOGGING;
   ```

3. Ensure that supplemental logging is enabled at the table level. To check, run the following commands for each relevant table:

   ```
   ALTER TABLE <schema>.<tablename> ADD SUPPLEMENTAL LOG DATA
   (ALL) COLUMNS;
   ```

Amazon Oracle RDS Database

To enable change tracking on an Amazon Oracle RDS database, supplemental logging must be enabled both at the database level and the table level.

1. To enable supplemental logging at the database level, run the following command:

   ```
   exec rdsadmin.rdsadmin_util.alter_supplemental_logging('ADD',
   'ALL');
   ```

2. To retain archived redo logs on your DB instance (e.g., for 24 hours), run the following command:

   ```
   exec rdsadmin.rdsadmin_util.set_configuration('archivelog
   retention hours', 24);
   ```

3. To enable supplemental logging at the table level, run the following command for each table:

   ```
   ALTER TABLE <schema>.<tablename> ADD SUPPLEMENTAL LOG DATA
   (ALL) COLUMNS;
   ```

Grants Required for Oracle Source Connector

The following grants and permissions are required to configure an Oracle database as a source connector:

- SELECT access on all tables to be replicated.
- SELECT access on V$ARCHIVED_LOG.

The database user must have the following security permissions:

- CREATE SESSION
- SELECT access on V_$LOGMNR_CONTENTS
- SELECT access on V_$LOGMNR_LOGS
- SELECT access on ANY TRANSACTION
- SELECT access on DBA_OBJECTS
- EXECUTE access on DBMS_LOGMNR

Run the following SQL grants for the required user (<user>):

```
GRANT SELECT ON V_$ARCHIVED_LOG TO <user>;

GRANT SELECT ON V_$LOGMNR_CONTENTS TO <user>;

GRANT EXECUTE ON DBMS_LOGMNR TO <user>;

GRANT SELECT ON V_$LOGMNR_LOGS TO <user>;

GRANT SELECT ANY TRANSACTION TO <user>;

GRANT SELECT ON DBA_OBJECTS TO <user>;

GRANT execute_catalog_role TO <user>;

GRANT LOGMINING TO <user>; -- Required for Oracle 12c and higher
```

- For Oracle 10g, create a public synonym for DBMS_LOGMNR:

  ```
  CREATE PUBLIC SYNONYM DBMS_LOGMNR FOR SYS.DBMS_LOGMNR;
  ```

  If using Oracle 19c with continuous log mining, grant additional permissions:

  ```
  GRANT SELECT ON V_$DATABASE TO <user>;

  GRANT SELECT ON V_$LOG TO <user>;

  GRANT SELECT ON V_$LOGFILE TO <user>;
  ```

- For Oracle Pluggable Databases (PDBs), ensure that all the above grants are run on the root database.

Verification of Oracle Source Setup

To verify that Oracle is correctly set up for change tracking, run the following commands and check the results.

Condition to be Checked	**SQL Query**	**Expected Result**
Is ArchiveLog mode enabled?	`SELECT log_mode FROM V$DATABASE;`	`ARCHIVELOG`
Is Supplemental Logging enabled at the database level?	`SELECT supplemental_log_data_min FROM V$DATABASE;`	`YES`
Is Supplemental Logging enabled at the table level?	`SELECT log_group_name, table_name, always, log_group_type FROM dba_log_groups;`	`<log group name>, <table name>, ALWAYS, ALL COLUMN LOGGING`

MySQL

Flow supports on-premises MySQL and Amazon RDS MySQL databases as source databases.

On-Premises MySQL Database

To enable change tracking on an On-Premises MySQL database, perform the following steps:

1. To enable binary logging, configure the following parameters in the MySQL configuration file:
    a. On MySQL for Windows: `my.ini`
    b. On MySQL for Linux: `my.cnf`
2. Add or modify the following parameters:

Parameter	**Value**
`server_id`	Any value starting from 1. Example: `server_id = 1`
`log_bin`	Path to the binary log file. Example: `log_bin = D:\MySQLLogs\BinLog`
`binlog_format`	Set to `row` for row-based logging. Example: `binlog_format = row`

`expire_logs_days`	Set to a non-zero value to avoid disk space issues. Example: `expire_logs_days = 4`
`binlog_checksum`	Set to `none` to disable checksums. Example: `binlog_checksum = none`
`binlog_row_image`	Set to `full` for capturing all changes. Example: `binlog_row_image = full`

Grants Required for On-Premises MySQL Source Connector

MySQL user must have the following privileges:

- **Replication Client**
- **Replication Slave**
- **Select privileges** on the source tables designated for replication.

Run the following commands to grant the necessary permissions to a MySQL user:

```
CREATE USER 'bflow_ingest_user' IDENTIFIED BY '*****';

GRANT SELECT, REPLICATION CLIENT, SHOW DATABASES ON *.* TO
'bflow_ingest_user' IDENTIFIED BY '*****';

GRANT SELECT, REPLICATION SLAVE, SHOW DATABASES ON *.* TO
'bflow_ingest_user' IDENTIFIED BY '*****';
```

Amazon RDS MySQL Database

To enable change tracking, perform the following steps within the AWS Management Console:

1. Create a new DB parameter group for the MySQL instance.
2. Set the following parameters in the newly created DB parameter group:

Parameter	**Value**
`binlog_format`	`binlog_format=row`
`binlog_checksum`	`binlog_checksum=none` or `binlog_checksum=CRC32`

3. Apply the new parameter group to the MySQL RDS DB instance.

Ensure that binary logging is enabled on the RDS instance for the change tracking setup to work properly.

Grants Required for Amazon RDS MySQL Source Connector

Ingest user must have the following privileges:

- **Replication Client**
- **Replication Slave**
- **Select privileges** on the source tables designated for replication.

Run the following commands to grant the necessary permissions to the MySQL RDS user:

```
CREATE USER 'bflow_ingest_user' IDENTIFIED BY '*****';

GRANT SELECT, REPLICATION CLIENT, SHOW DATABASES ON *.* TO
'bflow_ingest_user';

GRANT SELECT, REPLICATION SLAVE, SHOW DATABASES ON *.* TO
'bflow_ingest_user';
```

If the source DB is an Amazon RDS MySQL database, you must download the `mysqlbinlog.exe` utility. Add the directory path of `mysqlbinlog.exe` to the Windows **`'Environment variable' PATH`** on the Flow machine to ensure proper functioning.

PostgreSQL

To ensure integration between Flow and PostgreSQL database, perform the following steps:

1. Ensure the version of PostgreSQL is 9.4.x or later. The integration requires this version or higher for proper functioning.
2. To enable communication between Flow and PostgreSQL instance, configure access permissions in the `pg_hba.conf` file by adding the Flow machine's IP address with replication privileges. For example:

   ```
   host replication all 189.452.1.212/24 md5
   ```

   Replace `189.452.1.212/24` with the IP address of the Flow machine. Ensure that the IP address is allowed to access the database for replication purposes.

3. Set the following parameters in `postgresql.conf` file to ensure proper replication.
   a. Set `wal_level` to `logical`: Logical replication is required for Flow to track and replicate changes to the database.
   b. Set `max_replication_slots` to a value greater than `1`: `max_replication_slots` defines the maximum number of replication slots available for logical replication. This value must be set according to the number of concurrent tasks you plan to run. Each task requires a dedicated replication slot.

c. By default, `wal_sender_timeout` is set to 60 seconds. Connections that are inactive for longer than the specified timeout are terminated. To prevent connection termination during extended idle periods, SingleStore recommends setting this value to 0 to disable the timeout mechanism.

4. After configuring these parameters, restart PostgreSQL for the changes to take effect.
5. Once the PostgreSQL configuration is updated, create a logical replication slot for the database to sync with Flow. Run the following SQL command:

```
SELECT
pg_create_logical_replication_slot('bflow_replication_slot',
'test_decoding');
```

6. Grant Replication role to Flow user.
7. Grant **Read-only** access for all the replicating tables to Flow user.

SQL Server

To enable change tracking on a SQL Server database, perform the following steps:

1. Install `bcp` – a Microsoft Utility and the following drivers:
   a. VC++ 2017 64 bit
   b. ODBC drivers 18 64 bit
   c. SQL CMD version 15 64 bit

2. To enable change tracking at the database level, run the following SQL command for each database:

```
ALTER DATABASE <databasename>
SET CHANGE_TRACKING = ON
(CHANGE_RETENTION = 7 DAYS, AUTO_CLEANUP = ON);
```

3. To enable change tracking at the table level, run the following SQL command for each table:

```
ALTER TABLE <tablename>
ENABLE CHANGE_TRACKING
WITH (TRACK_COLUMNS_UPDATED = ON);
```

4. Ensure that change tracking is enabled for all databases and tables to be replicated.
5. The SingleStore Ingest database replication login user must have appropriate permissions to view change tracking information.
   a. To grant view permission to event viewer logs, run the following command:

```
GRANT VIEW SERVER STATE TO <AccountName>;
```

b. To allow the user to view change tracking information at the table level, run the following command:

```
GRANT VIEW CHANGE TRACKING ON <schema>.<tablename> TO
<User>;
```

Note: If you are configuring Ingest for a completely new SQL Server database, ensure that at least one transaction is performed on the database to generate the log sequence number. This is necessary for Flow to start tracking the changes.

Install Ingest

You can install Ingest in both Windows and non-Windows environments.

Install Ingest as a Windows Service

Note: Filenames may change from release to release.

1. Download and unzip the Ingest Build zip file: https://bryteflow.com/release/singlestore/ingest/402_b571/ingest.zip
2. Change the directory to the **ingest** folder and open `ingest.xml` file in a Notepad. Change the following lines as per your requirements:
   a. `<id>SingleConnect-Ingest</id>`
      This tag is used as an unique ID in Windows services. Each Ingest pipeline has its own ID.
   b. `<name>SingleConnect Ingest</name>`
      This tag is a display name in the Windows services. It must be descriptive and brief.
   c. `<description>SingleConnect Ingest</description>`
      This tag specifies a description of the Ingest service.
3. Install Ingest as a Windows service.
   a. Open a command prompt as an Administrator.
   b. Change the current directory to the **ingest** folder and run the `Install-Service.bat` file using the following command:

      ```
      Install-Service.bat
      ```

   c. The batch file runs commands to install the service. Once the command is completed, close the command prompt.
4. Ingest service must appear in the Windows services list like `SingleStore-Ingest`. The name must be similar as shown in step 2(b).
5. Ingest service starts automatically upon installation and must be running. You can start, stop, or restart the Ingest service similar to other Windows services.

6. Open Chrome and go to http://localhost:8081. If the application is installed on a server, go to `http://<server-ip-address>:8081` instead.
7. Navigate to **Settings > License**, and enter the license key.

Install Ingest in Non-Windows Environment

1. Download and unzip the Ingest Build zip file: https://bryteflow.com/release/singlestore/ingest/402_b571/ingest.zip

2. Run the following Java command from the directory where the files are unzipped, to start the Ingest. To include external libraries, skip to the next step.

   ```
   java -jar ingest.jar
   ```

3. To include external libraries, use the following command instead.
   a. For Windows:

      ```
      java -cp "other_lib.jar;ingest.jar" org.springframework.boot.loader.JarLauncher
      ```

   b. For Linux:

      ```
      java -cp "other_lib.jar:ingest.jar" org.springframework.boot.loader.JarLauncher
      ```

4. The default port for Ingest is 8081. You can change this port from the user interface (UI). The new port setting takes effect after the next run.

Install XL Ingest

You can install XL Ingest in both Windows and non-Windows environments.

Note: Filenames may change from release to release.

Install XL Ingest as a Windows Service

1. Download the XL Ingest Build zip file: https://bryteflow.com/release/singlestore/ixl/23_b1174/xlingest.zip

   Unzip the file into a directory at the same level as the Ingest directory.

2. Change the directory to the **xlingest** folder and open `xlingest.xml` file in a notepad. Change the following lines as per your requirements:
   a. `<id>SingleConnect-XL-Ingest</id>`
      This tag is used as an unique ID in Windows services. Each XL Ingest pipeline has its own ID.
   b. `<name>SingleConnect XL Ingest</name>`
      This tag is a display name in the Windows services. It must be descriptive and brief.

c. `<description>SingleConnect XL Ingest</description>`
   This tag specifies a description of the Ingest XL service.

3. Install XL Ingest as a Windows service.
   a. Open a command prompt as an Administrator.
   b. Navigate to **xlingest** folder and run the `Install-SingleConnect-XL-Ingest-Service.bat` file using the following command:

      ```
      Install-SingleConnect-XL-Ingest-Service.bat
      ```

   c. The batch file runs commands to install the service. Once the command is completed, close the command prompt.
4. Ingest service must appear in the Windows services list like `SingleStore-XL-Ingest`. The name must be similar as shown in step 2(b).
5. Ingest service starts automatically upon installation and must be running. You can start, stop, or restart the Ingest service like other Windows services.
6. Open Chrome and go to http://localhost:8084. If the application is installed on a server, go to `http://<server-ip-address>:8084` instead.
7. Navigate to **Configuration > License**, and enter the license key.

Install XL Ingest in Non-Windows Environment

1. Download the XL Ingest Build zip file:
   https://bryteflow.com/release/singlestore/ixl/23_b1174/xlingest.zip

   Unzip the file into a directory at the same level as the **ingest** directory.

2. Run the following Java command from the directory where the files are unzipped, to start the Ingest. To include external libraries, skip to the next step.

   ```
   java -jar xlingest.jar
   ```

3. To include external libraries, use the following command instead.
   a. For Windows:

      ```
      java -cp "other_lib.jar;xlingest.jar" cirus.Run
      ```

   b. For Linux:

      ```
      java -cp "other_lib.jar:xlingest.jar" cirus.Run
      ```

4. The default port for XL Ingest is 8084. You can change this port from the user interface (UI). The new port setting takes effect after the next run.

Apply the License Key

The Flow suite of products comes with a validity period specific to each customer's agreement. Ensure you get valid license keys to begin using the software. Each software product has its own unique product ID when set up individually.

Apply License Key in Ingest

To obtain a valid license, contact SingleStore Sales with the **Product ID**. Navigate to **Settings > Licence**. In **Licence Key**, enter the license key, and then select **Apply** to save the settings.

Apply License Key in XL Ingest

To obtain a valid license contact SingleStore Support with the **Product ID**. Navigate to **Configuration**. In **Licence Key**, enter the license key, and then select **Save** to save the settings.