



Data Engineering Guide

Version 2.7
2024-07-12

Data Engineering Guide

InterSystems Version 2.7 2024-07-12

Copyright © 2024 InterSystems Corporation

All rights reserved.

InterSystems®, HealthShare Care Community®, HealthShare Unified Care Record®, IntegratedML®, InterSystems Caché®, InterSystems Ensemble®, InterSystems HealthShare®, InterSystems IRIS®, and TrakCare are registered trademarks of InterSystems Corporation. HealthShare® CMS Solution Pack™ HealthShare® Health Connect Cloud™, InterSystems IRIS for Health™, InterSystems Supply Chain Orchestrator™, and InterSystems TotalView™ For Asset Management are trademarks of InterSystems Corporation. TrakCare is a registered trademark in Australia and the European Union.

All other brand or product names used herein are trademarks or registered trademarks of their respective companies or organizations.

This document contains trade secret and confidential information which is the property of InterSystems Corporation, One Memorial Drive, Cambridge, MA 02142, or its affiliates, and is furnished for the sole purpose of the operation and maintenance of the products of InterSystems Corporation. No part of this publication is to be used for any other purpose, and this publication is not to be reproduced, copied, disclosed, transmitted, stored in a retrieval system or translated into any human or computer language, in any form, by any means, in whole or in part, without the express prior written consent of InterSystems Corporation.

The copying, use and disposition of this document and the software programs described herein is prohibited except to the limited extent set forth in the standard software license agreement(s) of InterSystems Corporation covering such programs and related documentation. InterSystems Corporation makes no representations and warranties concerning such software programs other than those set forth in such standard software license agreement(s). In addition, the liability of InterSystems Corporation for any losses or damages relating to or arising out of the use of such software programs is limited in the manner set forth in such standard software license agreement(s).

THE FOREGOING IS A GENERAL SUMMARY OF THE RESTRICTIONS AND LIMITATIONS IMPOSED BY INTERSYSTEMS CORPORATION ON THE USE OF, AND LIABILITY ARISING FROM, ITS COMPUTER SOFTWARE. FOR COMPLETE INFORMATION REFERENCE SHOULD BE MADE TO THE STANDARD SOFTWARE LICENSE AGREEMENT(S) OF INTERSYSTEMS CORPORATION, COPIES OF WHICH WILL BE MADE AVAILABLE UPON REQUEST.

InterSystems Corporation disclaims responsibility for errors which may appear in this document, and it reserves the right, in its sole discretion and without notice, to make substitutions and modifications in the products and practices described in this document.

For Support questions about any InterSystems products, contact:

InterSystems Worldwide Response Center (WRC)

Tel: +1-617-621-0700

Tel: +44 (0) 844 854 2917

Email: support@InterSystems.com

Table of Contents

1 Welcome, Data Engineers	1
1.1 Data Sources and Schemas	1
1.2 Recipes	1
1.3 Analytics Cubes	2
1.4 Snapshots	2
1.5 Scheduling	2
1.6 See Also	3
2 Importing Schemas	5
2.1 Adding Tables from a JDBC Data Source	5
2.2 Adding an SQL Query from a JDBC Data Source	6
2.3 Adding a File-Based Schema	6
2.4 Filter Options When Importing Schemas	7
2.5 See Also	8
3 Schema Evolution	9
3.1 How the System Manages Changes to Schemas	9
3.2 See Also	9
4 Editing and Managing Schemas	11
4.1 Viewing the Data Catalog	11
4.2 Displaying a Schema	12
4.3 Editing a Schema	12
4.4 Schema Options	12
4.5 Editing Schema Fields	15
4.6 Publishing a Schema	15
4.7 Reimporting a Schema	15
4.8 Deleting a Schema	16
4.9 See Also	16
5 Defining and Managing Categories	17
5.1 Defining a Category	17
5.2 Renaming a Category	17
5.3 Deleting a Category	18
5.4 See Also	18
6 Introduction to Recipes and Staging Tables	19
6.1 Structure of a Recipe	19
6.2 Staging Tables	20
6.3 See Also	20
7 Defining Recipes	21
7.1 Creating a Recipe	21
7.2 Adding a Staging Activity	21
7.3 Adding a Transformation Activity	23
7.4 Transformation Functions	23
7.5 Adding a Validation Activity	24
7.6 Validation Functions	25
7.7 Adding a Reconciliation Activity	26
7.8 Viewing the Staging Table Details	28

7.9 Adding a Promotion Activity	28
7.10 Adding a Custom Activity	30
7.11 Step Mode	30
7.12 See Also	31
8 Editing and Managing Recipes	33
8.1 Recipe Changes	33
8.2 Viewing the Recipes	33
8.3 Editing a Recipe	34
8.4 Publishing an Activity	34
8.5 Disabling an Activity	35
8.6 Enabling an Activity	35
8.7 Deleting an Activity	35
8.8 Renaming a Recipe	36
8.9 Moving a Recipe Into or Out of a Group	36
8.10 Copying a Recipe	37
8.11 Disabling a Recipe	37
8.12 Enabling a Recipe	37
8.13 Purging Old Recipe Data	38
8.14 Deleting a Recipe	38
8.15 Recovering a Recipe	38
8.16 Permanently Deleting a Recipe	39
8.17 See Also	39
9 Managing Recipe Groups	41
9.1 Creating a Recipe Group	41
9.2 Renaming a Recipe Group	41
9.3 Deleting a Recipe Group	42
9.4 Recovering a Deleted Recipe Group	42
9.5 See Also	42
10 Defining Snapshots	43
10.1 Uses of Snapshots	43
10.2 Snapshot Tables	43
10.3 Defining a Snapshot	43
10.4 See Also	45
11 Editing and Managing Snapshots	47
11.1 Viewing the Snapshots Dashboard	47
11.2 Editing a Snapshot	47
11.3 Retagging a Snapshot Run	48
11.4 Disabling a Snapshot	48
11.5 Enabling a Snapshot	48
11.6 Deleting a Snapshot	49
11.7 See Also	49
12 Managing Table Indices	51
12.1 Adding an Index to a Staging Table	51
12.2 Adding an Index to a Snapshot	51
12.3 Index Build Pending Message	52
12.4 See Also	52
13 Defining Entities and Calendars	53
13.1 Defining an Entity	53

13.2 Defining a Child Entity	54
13.3 Editing an Entity	54
13.4 Editing a Holiday Calendar	55
13.5 Deleting an Entity	56
13.6 See Also	56
14 Scheduling and Running Tasks	57
14.1 Scheduling a Task	57
14.2 Managing Task Dependencies	58
14.3 Modifying a Task	59
14.4 Running a Task Manually	59
14.5 Aborting a Task	60
14.6 Deleting a Task	60
14.7 See Also	60
15 Handling Task Errors (Using the Workflow Inbox)	61
15.1 When the System Encounters an Error	61
15.2 First Steps	61
15.3 Viewing the Workflow Inbox	62
15.4 Addressing a Task	62
15.5 Relinquishing a Task	63
15.6 See Also	63
16 Viewing Run History	65
16.1 Run History of a Recipe	65
16.2 Recipe Run Details and Downloadable Reports	65
16.3 Run History of a Snapshot	66
16.4 Run History from the Business Scheduler	66
16.5 Task Status	67
16.6 See Also	67
17 Filtering and Customizing the Business Scheduler	69
17.1 Default Display	69
17.2 Filtering the Display	70
17.3 Customizing the Columns	70
17.4 See Also	70
18 Using the File Manager	71
18.1 Standard Subdirectories for a File Source	71
18.2 Viewing the Directory Contents for a File Source	71
18.3 Uploading Files	72
18.4 Deleting a File	73
18.5 See Also	73
19 Exporting and Importing Configurations	75
19.1 Exporting Configurations	75
19.2 Importing Configurations	75
19.3 See Also	76

1

Welcome, Data Engineers

As a data engineer for your [InterSystems® Data Fabric Studio™](#) solution, your task is to set up the *data pipeline*, a term that refers to all the infrastructure needed to make data available within the solution. This page provides an overview of that infrastructure, from start to finish.

When data is available within the system, you and [data analysts](#) can create analysis cubes and reports based on those tables.

1.1 Data Sources and Schemas

In the product, a *data source* is a named configuration item that provides all the information needed to retrieve data from an external source. A classic data source is a database, but another possibility is a financial system that provides an API via which an authenticated user can retrieve data. Yet another option is delimited files, either retrieved from a cloud location or pushed to the local file system. In all cases, the [data source](#) definition, usually configured by an administrator, contains all the information needed to retrieve structural information as well as data.

A data source can provide multiple kinds of data, each with a specific structure. The classic case is a database that contains many tables, where each table has a specific structure. Similarly, a specific API call returns data in a specific format. In the product, each of these structures is a *schema*, and the goal is to create a Data Catalog that consists of the schemas that are relevant to the business (and that are ultimately needed for reports and analysis). As a data engineer, your task is to import schemas from the data sources into the Data Catalog and then refine them and categorize them in ways that are useful to your organization.

To get started defining schemas, see [Importing Schemas](#).

1.2 Recipes

Recipes describe how to load data from external sources into Data Fabric Studio or into external tables. Any recipe consists of some or all of the following steps, in order:

1. Staging activities, each of which loads data into a staging table.
2. Transformation activities, which can clean data in various ways.

Transformation activities update the staging table, either by adding new fields or overwriting existing fields. You can directly examine the staging table at any time.

3. Validation activities, which can compare data to desired ranges, as an example.

4. Reconciliation activities, in which you specify comparisons that define a valid reconciliation of the data.

Each validation and reconciliation activity writes a report file with any errors. For either kind of activity, an error can either halt processing or simply result in a warning message, as you choose. In all cases, an appropriate user role is alerted via the [workflow](#) module so that the data can be examined and corrected if needed, and the recipe can be rerun.

5. Data promotion activities, which uses SQL to update a final table, based on the contents of one or more staging tables. The final table can be in the native database or can be in an external database.

A recipe can also include custom steps at any stage in the processing.

To get started with this task, see [Defining Recipes](#).

1.3 Analytics Cubes

Data Fabric Studio includes InterSystems IRIS® Adaptive Analytics, a Business Intelligence tool powered by co-development with [AtScale](#). This means that data engineers and [data analysts](#) can define cubes based on those tables, and then use those cubes for analytics. See [Defining Cubes](#).

1.4 Snapshots

Data Fabric Studio provides an additional mechanism to support your data needs: snapshots. With snapshots, you can easily save data for later inspection by regulators; the snapshot can pull data from multiple tables as needed, writing them to a snapshot table, and the system applies a tag to the records. The product automatically stores all snapshot runs; that is, a new snapshot run does not overwrite a previous snapshot run.

When you have run a snapshot multiple times, applying a different tag each time, you can examine how that data changes over time. In particular, you can build a cube on the snapshot data, using the tag values as a dimension.

See [Defining Snapshots](#).

1.5 Scheduling

The Business Scheduler provides an easy way to schedule the running of tasks: running recipes, building cubes, and performing snapshots. See [Scheduling and Running Tasks](#).

In the initial phases of implementation, you can configure tasks to be run manually (from the Business Scheduler).

Later, when you want the tasks to be executed on a schedule, it is necessary to define the applicable calendar information and to manage dependencies among tasks. This works as follows:

- To define calendar information, you define *entities*, each which has its own calendar. To simplify scheduling, the product supports a hierarchical system of entities, each of which can have its own business calendar but can inherit calendar details from its parent. See [Defining Entities and Calendars](#).
- To specify dependencies among tasks, you apply a tag to a scheduled task and define dependency expressions in other tasks, referring to that tag. Then a task can be run (on a given day) only after the dependencies are fulfilled. See [Managing Task Dependencies](#).

1.6 See Also

- [Welcome, Data Analysts](#)
- [About Your Solution: What Is Not Documented](#)

2




Importing Schemas

This page describes how to import the *schemas* that make up the [Data Catalog](#). Each of these schemas (not to be confused with SQL schemas) is associated with a specific external [data source](#) and describes the structure of a single data element such as a table or a delimited file provided by that data source.


After importing a schema, it is generally necessary to [edit](#) the details such as the data extraction strategy and field types.

2.1 Adding Tables from a JDBC Data Source


To add one or more tables from a JDBC data source:

1. Click the Data Catalog  icon in the application menu.
2. Click **Data Schema Importer**.
3. Select a JDBC data source from the **Data Source** dropdown menu.
4. Select an SQL schema (group of tables) from the **Schema** dropdown list.
The section **Available Items from JDBC Source** then displays the tables in the selected SQL schema.
5. Add tables to be imported in either of two ways:
 - Click the check box next to each table name and then click the Add to Imports  button above the table.
 - Click the Add to Imports  button in the applicable row or rows.

The table or tables are then moved to the section **Items Selected For Import**; the system has not yet imported their metadata.


6. Optionally select a different item from the **Schema** dropdown list and repeat these steps as needed. This process adds to the list in **Items Selected For Import**; the system has not yet imported their metadata.
7. Optionally, to remove a table from the list in **Items Selected For Import**, click the Delete  icon in the applicable row.
8. To import the metadata for the selected tables, click **Import**.

The system then imports metadata for all the selected tables and displays the **Results of Last Import** tab, which lists all the tables whose metadata it just imported.

- From the **Results of Last Import** tab, optionally start refining any of the imported schemas. To do so, click the Inspect  icon in the applicable row. Then edit that item as described in [Editing Schemas](#).

2.2 Adding an SQL Query from a JDBC Data Source

Instead of (or in addition to) adding [tables](#) from a JDBC data source, you can directly use a custom SQL SELECT query. To do so:

- Click the Data Catalog  icon in the application menu.
- Click **Data Schema Importer**.
- Select a JDBC data source from the **Data Source** dropdown menu.
- Click the **Query** tab.
- Type an SQL SELECT statement into the box.
- Optionally modify **Row Count**, which controls the number of rows that are sampled and displayed.
- Click **Run Query**.

The page then displays the query results, which you can use to verify that the query is as expected.

- To add a schema to the catalog based on this query, click **Save to Catalog**.

The system then displays a dialog box.

- For **Name**, type a short, unique name for the new schema. You cannot change this name later.
- For **Description**, type an optional description of the new schema.

The system then generates metadata for the query and displays the new schema, which you can now edit as described in [Editing Schemas](#).

2.3 Adding a File-Based Schema

A file-based data source is a UTF-8 encoded file with one record per line, where each line follows a convention that uses a specific delimiter between fields (typically a comma or a tab). This convention implicitly defines a schema. To define this schema within the product, you need to upload and then import a sample file and specify the delimiter and the field names. (Excel files are a special case of a file-based data source in which you do not need to specify the delimiter that separates fields.)

[FileDir](#), [ExcelSingleFileDir](#), and [S3Delimited](#) data sources are all file-based data sources.


To add a file-based schema:

- Obtain a sample file for the schema, applicable to a specific file-based data source.



The sample file can consist of only one line. Also, the sample file must have a header row—an initial line that contains the names of the fields. (In other words, the sample file may or may not contain any actual data.)

- Rename the sample file so it has an appropriate and useful short filename, because this short filename becomes the schema name within the system (with punctuation characters removed), and it cannot be edited.


For example, if you load a sample file named `sampledata.csv`, that becomes a schema named `sampledatacsv`.

3. Use the [File Manager](#) to upload this file to the Samples directory used by the chosen file-based data source.
4. Click the Data Catalog  icon in the application menu.
5. Click **Data Schema Importer**.
6. Select the chosen file-based data source.

The section **Available Items from Source** lists the files in the Samples [subdirectory](#) for this data source.

7. Add files to be imported in either of two ways:
 - Click the check box next to each filename and then click the Add to Imports  button above the table.
 - Click the Add to Imports  button in the applicable row or rows.

The filename or filenames are then moved to the section **Items Selected For Import**; the system has not yet imported their metadata.

8. If the right side of the page contains a long list of files to import, you may want to double check the list. In doing so, if you want to filter this display, you can type into the **Members Selected for Import** filter. This option affects what this page *displays*, but does not affect the import of metadata.
9. Optionally, to remove a file from this list, click the Delete  icon in the applicable row.
10. To import the selected files, click **Import**.

The system then imports all the selected files and generates metadata from them.

The page then displays the **Results of Last Import** tab, which lists all the files it imported.

2.4 Filter Options When Importing Schemas

Depending on the type of data source, there may be a large number of schemas to choose among. The import page provides options to help you filter the display. You can filter and sort the items listed in the **Available Items From Source** section on the left, as well as the items listed in the **Items Selected For Import** on the right. In both areas, you can do the following:

- To show only names containing a specific string, type that string into the filter above the **Item Name** column.
- To display based on whether they are in the catalog, click the dropdown above the **In Catalog** column and select either or both of the following check boxes as needed:
 - **Yes**—display the items that are currently in the catalog.
 - **No**—display the items that are not currently in catalog.
- You can click the **Item Name** and **Item Name** column headers to sort those columns.

Important: These filters do not affect what is imported when you click **Import**. The only change is to the display of the items on this page.

2.5 See Also

- [Defining Data Sources](#)
- [Using the File Manager](#)
- [Editing and Managing Schemas](#)
- [Defining Categories](#)

3

Schema Evolution

This page explains how InterSystems® Data Fabric Studio™ accommodates changes while also preserving your data, a process known as *schema evolution*.

3.1 How the System Manages Changes to Schemas

Schema evolution works as follows:

1. When you [import](#) a schema and first edit it, the schema is a draft and cannot be used by recipes.
2. When you [publish](#) a schema, it can then be used by recipes—specifically in a staging activity (which loads data into the system).
3. When you define a [recipe](#), each of its activities is initially a draft.
4. When you publish a staging activity of a recipe, the system then generates the table that will store the data to be loaded.
5. When you [edit](#) or [reimport](#) a schema, you are implicitly creating a draft that is not yet used by anything. The existing schema is unchanged; existing data is also unchanged.

While a schema is in draft state, you have the option of deleting the draft (and reverting to the previous published version).

6. When you [republish a schema](#), if the change affects the structure needed to contain the data, the system automatically increments the version number of the schema and automatically generates a new draft of any staging activity that is affected by this change. The new staging activity draft is not automatically published, and the previous staging activity is unchanged.
7. When you [republish a staging activity](#), the system generates a new table as needed, considering the structural change that has occurred. This leaves the old table (and its data) unchanged.

Similarly, changes to earlier parts of a recipe can affect later activities in a recipe, requiring changes in those later activities. The system detects these changes, automatically generates draft activities where appropriate, and notifies you of the drafts.

3.2 See Also

- [Editing and Managing Schemas](#)
- [Recipes and Staging Tables](#)

- [Defining Recipes](#)

4

Editing and Managing Schemas

In InterSystems® Data Fabric Studio™, a schema describes the structure of a single data element such as a table or a file using a specific file format. This page describes how to edit and manage schemas; another page provides information on [importing](#) them.


4.1 Viewing the Data Catalog

To view the Data Catalog and the schemas currently contained in it:

1. Click the Data Catalog  icon in the application menu.
2. Click **Data Catalog Browser**.


This page enables you to see all available schemas, either as a table or as a set of rectangular cards.

For each schema, the page displays the name of the data source that contains this schema, the actual name of the schemas, and any [categories](#) to which it belongs. If you click **Card View**, the cards also display the descriptions.


3. Optionally filter the display by using the options at the top of the page:
 - **Schema Definition Item Name**—Filters the display to include only the schemas whose short names contain the given string. After typing into this field, click the Find  icon.
 - **Data Source**—Filters the display to include only schemas from a specific [data source](#).
 - **Category**—Filters the display to include only schemas labelled with a specific [category](#). Note that schemas can be labelled with any number of categories, including no categories at all.
 - **Recipe**—Filters the display to include only schemas used in a specific [recipe](#).
 - **Drafts** —If selected, displays only the schemas that are [drafts](#).
 - **Reset** —Clears all the filter options.

Click any schema in order to view it or make changes.

4.2 Displaying a Schema

If you have just imported the schema and you are viewing the **Results of Last Import** tab, click the Inspect  icon in the row for a schema.

Otherwise, to display a schema:

1. Click the Data Catalog  icon in the application menu.
2. Click **Data Catalog Browser**.
3. Click the schema name.



If there is a draft of the schema, you can choose to display the published version or display the draft (which is shown as a separate item as in the following example).



4.3 Editing a Schema

To edit a schema, first [display](#) it. If the schema has been published, click **Edit Schema** in the upper right to make the schema editable. (If there is a draft of the schema and you are viewing the published version, this button opens the draft version of the schema.)

Then make any of the following changes:

- To add or edit a description, click the Edit  icon next to **Description** and then make edits.
- To add a category, click **Add Category...** and then click a category.
- To remove a category, click the X icon for the category.
- To modify the details of how the data source is handled, click the Edit  icon in that right box. This displays a dialog box where you can make changes. The details depend on the kind of schema; see [Schema Details](#).
When you are done, click **Save** to save these changes or **Cancel** to discard them.
- To make changes to the schema fields, see the [next section](#).

When you are done with the schema, click **Save Draft** to save these changes or **Delete Draft** to discard them (and return to the previous schema definition).

Note that if you leave this page without saving your changes, the system automatically saves your changes as a draft.

4.4 Schema Options

This section provides reference information on the options that control how a schema (and its data source) are interpreted and used.

Extraction Strategy

Applies to all schema types. Specifies which records to use from the data source.

For SQL-based schemas (tables and custom queries), the options are as follows:

- **General Table**—Retrieves all new records and any changed records.
A *new record* is a record with a new primary key value. A *changed record* is a record with an existing primary key value but a change to one or more other fields.
- **Append only transactional table with a numeric sequential key**—Retrieves all new records since the last sequential primary key.
If you specify this, also specify **Extraction Strategy Field**, which must be the name of the field that contains the sequential primary key.
- **Transactional table with a changed timestamp field**—Retrieves all records changed since the last timestamp.
If you specify this, also specify **Extraction Strategy Field**, which must be the name of the field that contains the timestamp.
- **Simple load**—Retrieves all records, whether or not the system has them already.
- **Change data capture**

For file-based schemas, the options are **General Table** and **Simple Load**.

UsePrimaryKeyPseudocolumn

Applies to table-based schemas. Specifies whether to use a pseudocolumn in this table as the primary key. If you select this option, also specify **PrimaryKeyPseudocolumnName**, which must be the name of the pseudocolumn.

Does not apply to a schema based on a custom query.

CreateForeignTable

Applies to SQL-based schemas (tables and custom queries). Projects this table or query as a foreign table, which is read-only and which does not copy the data into Data Fabric Studio.

If you select this option, you can override the default schema used for foreign tables from this data source. To do so, select **ForeignTableLocalSchemaOverride** and then specify a schema name for **ForeignTableLocalSchema**. (An additional override is available for individual tables, if needed.)

Delimiter

Applies to file-based schemas. Specifies the delimiter that separates fields in the file.

When loading delimited files, you must define and follow specific conventions so that they can be read appropriately. The first consideration is the delimiter that separates the fields, controlled by the **Delimiter** setting.

If the character specified by **Delimiter** is also used *within* any field, you must wrap the field in quotation marks (single or double, as you choose—or possibly some other character). To control which character is used to wrap a field, you specify the **QuoteChar** option, whose default is the double quote character ". If the character specified by **QuoteChar** is used *within* any field that is already wrapped in quotation marks, see the **DoubleQuote** and **EscapeChar** settings.

FileNamePattern

Applies to file-based schemas. Specifies which files are read and used as input, for this schema. That is, input files whose names conform to this pattern are processed, and other files are ignored. You can specify any legal filename string or a filename wildcard expression, consistent with the operating system. The initial pattern is simply the name of the file that you initially loaded.

FileNamePatternMatching

Applies to file-based schemas. Specifies the form of pattern matching to use (for **FileNamePattern**).

CommentLinePrefix

Applies to file-based schemas. Specifies the prefix that indicates the start of a comment line in this file. A comment line is ignored.

DoubleQuote

Applies to file-based schemas. Specifies how to interpret a pair of quotation marks contained within a field that is wrapped in quotation marks. (Here the term *quotation marks* refers to the character specified by the **QuoteChar** option.)

If this option is **Yes**, then in the stated context, a pair of quotation marks is interpreted as a single quotation mark.

For example, suppose that **Delimiter** is a comma, **QuoteChar** is `"`, and **DoubleQuote** as **Yes**. Then the value `"here is a "value" to load"` is loaded as `here is a "value" to load`

EscapeChar

Applies to file-based schemas. Specifies the character used to start an escape sequence, so that you can include a quotation mark within a field that is wrapped in quotation marks. *Applies to file-based schemas.* Specifies how to interpret a pair of quotation marks contained within a field that is wrapped in quotation marks. (Here the term *quotation marks* refers to the character specified by the **QuoteChar** option.)

For example, suppose that **QuoteChar** is `"` and **EscapeChar** is a backslash `\`. Then the value `"here is a \"value\" to load"` is loaded as `here is a "value" to load`

LineTerminator

Applies to file-based schemas. Specifies the character or sequence of characters that indicate the end of a line in this file.

QuoteChar

Applies to file-based schemas. Specifies the character used to wrap any field that contains the field delimiter character. For example, if fields are delimited by a comma, then if a field contains a comma, that field must be wrapped in a quotation mark, as specified by **QuoteChar**.

If the character specified by **QuoteChar** is used *within* any field that is already wrapped in quotation marks, see the **DoubleQuote** and **EscapeChar** settings.

SkipInitialSpace


Applies to file-based schemas. Specifies how to handle space characters that appear within a field after the field delimiter. If this option is **Yes**, initial space characters are trimmed when the data is loaded. Note that you can also define a [transformation](#) that trims whitespace on a field-by-field basis.

Header Row

Applies to file-based schemas. Specifies whether the files have a header row. Select **Yes** or **No**. The header row is ignored when the data is loaded.

4.5 Editing Schema Fields

For any kind of data source, when you edit the schema, the bottom part of the page displays a table labeled **Schema Fields**. Here you can modify the schema fields as follows:

- To specify the primary key for this schema, click the **Primary Key** check box for all the fields that make up the primary key.
- To modify other details for a specific field, click the Edit Attributes  icon for that field. This displays a dialog box where you can edit the following:
 - **Data Type**—Specifies the data type for this field.
 - **Description**—Specifies an optional description for the field.
 - **Required**—Specifies whether this field is permitted to be null. If a field is required, an error is thrown if the value is null.
 - **Default Value**—Specifies the value to use if this field is null.
 - **Data Format** (for date, time, and timestamp fields)—Select the format that data is expected to be in for this field.
 - **Min Value** (for double and integer fields)—Specifies the minimum allowed value of this field.
 - **Max Value** (for double and integer fields)—Specifies the maximum allowed value of this field.
 - **Scale Value** (for double fields)—Specifies the number of digits to display after the decimal point. This has no effect on how data is stored.
 - **Length Value** (for string fields)—Specifies the maximum number of characters allowed in this field.

When you are done with this dialog box, click **Save** to save these changes or **Cancel** to discard them.

4.6 Publishing a Schema

A schema cannot be used in [recipes](#) unless it has been published. To publish a schema:

1. [Display](#) it.
2. Click **Publish Schema** in the upper right.

4.7 Reimporting a Schema

To reimport a schema:

1. [Display](#) it.
2. Click **Re-import Schema** in the upper right. This change automatically creates a new draft of the schema.
3. [Edit](#) the schema again as needed.

If the schema has not yet been [published](#), you cannot reimport it. You can, however, delete it and then import it again.

4.8 Deleting a Schema

To delete a schema:

1. [Display](#) it.
2. Click **Delete** in the upper right.
3. Click **Delete** to confirm.

You cannot delete a schema that is used in a [recipe](#); if you attempt to do so, you are notified about any recipe activities that depend on this schema. You can then go edit the recipe to remove the dependency.

4.9 See Also

- [Importing Schemas](#)
- [Defining Categories](#)
- [Defining Recipes](#)


5

Defining and Managing Categories

Categories enable you to define groupings of schemas in the [Data Catalog](#); each schema can belong to any number of categories (including none at all). Then when you search for a schema, you can use the categories as filters.



5.1 Defining a Category

To create a category:

1. Click the Data Catalog  icon in the application menu.
2. Click **Categories**.
The system displays any existing categories.
3. Click **New Category** in the upper right.
4. For **Name**, enter a short, unique name.
Category names cannot differ only by case. For example, you cannot have a category named `sample` and another category named `SAMPLE`.
5. Click **Submit**.

5.2 Renaming a Category



To rename a category:

1. Click the Data Catalog  icon in the application menu.
2. Click **Categories**.
The system displays any existing categories.
3. Click the Edit  icon in the row for the category.
4. For **Name**, enter a new name.
5. Click **Submit**.

The category name is automatically updated in all places where it is used.

5.3 Deleting a Category

To delete a category:

1. Make sure that no schema belongs to this category. To do this, display the [Data Catalog](#), filter to this category, and edit the schemas as necessary.
2. Click the Data Catalog  icon in the application menu.
3. Click **Categories**.
The system displays any existing categories.
4. Click the Delete  icon in the row for the category.
5. Click **Delete** to confirm this action.

5.4 See Also

- [Importing Schemas](#)
- [Editing Schemas](#)

6

Introduction to Recipes and Staging Tables

Recipes describe how to load data from external sources into InterSystems® Data Fabric Studio™ or into external tables. When you define a recipe that loads data into InterSystems® Data Fabric Studio™, the system generates staging tables and generates new versions of the tables in response to published changes in the data and in the recipe. This page provides an overview of a recipe and the generated staging tables.

6.1 Structure of a Recipe

Any recipe consists of some or all of the following steps, in order:

1. **Staging** activities. For each staging activity, you select a data source and then the schemas of interest from that data source.

When the recipe is run, the data for each schema is loaded into one or more generated **staging tables**. The system automatically uses the extraction strategy specified in the schema. In the recipe, you choose which fields to copy.

2. **Transformation** activities, which transform fields in the staging tables. You choose how to transform the value, and you specify whether to replace the existing value or save the transformed value into a new field that you specify (in the same table as the original value).
3. **Validation** activities, which examine the values in the staging table or tables. For each fields that you choose to validate, you specify the validation function.

The results of the validation checks are written to a validation report file. Validation errors can halt the recipe or can simply issue warnings, as you choose; in case of any validation error, the system generates a workflow task for the assigned role (specified when the recipe is **scheduled**).

4. **Reconciliation** activities, which compare values in different staging tables from the same run of the recipe. You specify how to match up records and you specify which fields to compare. Then for each comparison, you choose a comparison option.

The results are written to a reconciliation report file. Reconciliation errors can halt the recipe or can simply issue warnings, as you choose; in case of any reconciliation error, the system generates a workflow task for the assigned role (specified when the recipe is **scheduled**).

5. **Data promotion** activities. For each data promotion activity, you define a set of steps, each of which uses SQL to update a final table, based on the contents of the staging table or tables. This final table can be in the native database or can be in an external database.

A recipe can also include custom steps at each stage in the processing.

6.2 Staging Tables

As a preliminary step, a recipe involves loading data into generated staging tables. The names of these tables are generated, and the table name format is as follows:

`Staging_recipeshort_stagingshort_vversion.schemaName`

Where *recipeShort* is the **Recipe Short Name**, *stagingshort* is the short name of the [staging](#) activity, *version* is the version of the recipe, and *schemaName* is the name of the schema as seen in the [Data Catalog](#). (Note that `Staging` is a default part of this name but is [configurable](#).)

Any staging table includes the following fields:

- Fields maintained and used internally by the system: `%IRISRowID` and `%BatchId`. The `%BatchId` field contains the ID of the batch in which the given record was updated; you use this to identify the records to use in the data promotion activity.
- A set of fields whose names match the field names from the data that is loaded into that table; these fields are automatically added as needed.
- Any additional fields defined within the [transformation](#) activities. These field names start with the prefix `%T_`.

As you create a recipe, you may find it helpful to [view](#) the structure of the staging table, particularly when you define the promotion activities.

6.3 See Also

- [Schema Evolution](#)
- [Defining Recipes](#)
- [Editing and Managing Recipes](#)
- [Scheduling and Running Tasks](#)
- [Viewing Run History](#) (includes information on viewing dropped records, validation reports, and reconciliation reports)


7

Defining Recipes

This page discusses creating recipes; other pages discuss the details of [editing and managing recipes](#) and of [managing recipe groups](#).

7.1 Creating a Recipe


To define a recipe:

1. Click the Recipes  icon in the application menu.
2. Click **New Recipe**.
3. For **Recipe Name**, enter a descriptive name to display in the Recipes Dashboard.
4. For **Recipe Short Name**, enter a short name to be used within the names of the [generated tables](#) used by this recipe.
5. For **Recipe Group**, optionally select the name of a [recipe group](#).
You can move this recipe later to another group or remove it from the group.
6. Click **Submit**.

The system then displays a page where you can add activities to the recipe.

7.2 Adding a Staging Activity

To add a staging activity to a recipe:

1. Display the recipe:
 - a. Click the Recipes  icon in the application menu.
 - b. If the recipe is in a group, expand that group.
 - c. Click the recipe name.
The recipe is then displayed on the right side of the page.
2. Click **Add Staging Activity**.

3. For **Name**, enter a descriptive name to display within the recipe.
4. For **Short Name**, enter a short name to be used within the names of the [generated tables](#) used by this recipe.

If you are going to use only one data source in this recipe, you could use a generic short name like `Load`, but if you are going to use multiple data sources, it may be helpful for the short names to indicate the data source.

5. For **Data Source**, select a [data source](#).
6. Click **Submit**.

The system then displays a page where you can edit the details of how to use this data source. This page lists all the schemas that belong to this data source, and each schema represents a data asset such as a table, an specific API call, or a specific format of file. You can load data from any number of these items, within this staging activity.

7. For each schema of interest, click the schema name in the left. The right side of the page then displays information about the schema, including (if applicable) information about the last time data was loaded from the associated data source.
8. Then specify how to use this data source. First, optionally select the following check boxes:
 - **Stage on Target Table**—Enables you to copy data to the table name of your choice. The table is generated for you. If you select this, you are prompted for a full table name.
 - **Use SQLQuickload**—Enables the recipe to perform a bulk load for this step. If you select this, also specify the following options:
 - **Action On Dropped Records**—Specifies the action to take when the bulk load encounters more than the specified number of dropped records. The **Use SQLQuickload** option drops a record if that record cannot be inserted, which can occur if the record is invalid in some way (such as not having a required field or having a value in an incorrect format). For **Action On Dropped Records**, select one of the following, to specify how the system should behave if the maximum number of dropped records is reached:
 - **Abort**—The recipe is automatically aborted, without any workflow intervention.
 - **WorkflowStrict**—The recipe is stopped and a workflow task is generated. The workflow task provides the options **Abort** and **Retry**.
 - **Workflow**—The recipe is stopped and a workflow task is generated. The workflow task provides the options **Abort**, **Retry**, and **Ignore and Proceed**.
 - **Proceed**—The recipe continues.
 - **Max Dropped Records**—Specifies the maximum number of dropped records before the specified action is triggered.
9. Choose the fields to load. To do so, either click the check box at the top of the column of check boxes or click the check box for each field of interest.
10. When you are temporarily done with this activity, click **Save Draft**. Or if you are done defining the activity, click **Publish Activity**.

When you save the draft for the first time, the **Staging Table** field is updated to show the name of the [staging table](#). The table is not generated until you publish the activity. If you want to add additional activities that use the staging table, it is necessary to publish the staging activity.

11. Click the **Recipe** link in the upper left to return to the recipe.

7.3 Adding a Transformation Activity

To add a transformation activity to a recipe:

1. Display the recipe.
2. Click **Add Transformation Activity**.
3. For **Name**, enter a descriptive name to display within the recipe.
4. For **Target Source**, select a schema; the list includes all schemas used in *published* [staging activities](#) within the same recipe.

The system then displays the fields in the associated staging table, with options to specify transformations for any of the fields.

5. For each field where you want to apply a transformation, select the field name.

The system then displays options on the right side of the page. Make edits described below and then click **Apply**.

- **Function Type**—See [Transformation Functions](#).
- **Overwrite**—Select this if you want the new value to replace the old value. In this case, the system will not generate a new field in the staging table.

Otherwise, leave **Overwrite** cleared and specify a new field name (see below).

- **Transformation Name**—Specify a short name for this field transformation.
 - **Result Field Name**—Specify a name for the new field that will store this transformed value. In this case, the system will automatically add a field to the staging table; the actual name of the new field is %T_*newname* where *newname* is the name you specified.
 - **Result Field Type**—Select the type for the new result field.
 - Other options—See [Transformation Functions](#).
6. When you are done editing this activity, click **Save Draft**. Or click **Publish Activity** so that the recipe will use it.
 7. Click the **Recipe** link in the upper left to return to the recipe.

7.4 Transformation Functions

This section provides reference information on the [transformation](#) functions.

Code Table Conversion

Available for string fields. This function transforms the data by using a code table, which provides a set of key/value pairs. If the value being converted matches a key in the given code table, that value is converted to the value associated with that key. Also specify:

- **Code Table Name**—Specify the name of a table that contains key/value data.
- **Code Table Key Column**—Specify the name of the field in that table that contains the keys.
- **Code Table Value Column**—Specify the name of the field in that table that contains the associated values.

Custom SQL Expression

Available for string fields. This function transforms the data by using an SQL expression. Also specify **SQL Expression** as an SQL expression that returns a single value, such as an expression that combines field names and SQL operators. You must delimit any field name that is an SQL reserved word.

Lowercase String

Available for string fields. This function converts the string to lowercase.

Replace Substring

Available for string fields. This function performs a substring replacement. Also specify:

- **Old Substring**— Type the text to be replaced.
- **New Substring**— Type the replacement text.

Trim Whitespace

Available for string fields. This function trims whitespace (leading, trailing, or both). Select **Trim Leading Whitespace**, **Trim Trailing Whitespace**, or both.

Trim Characters

Available for string fields. This function trims a specified set of characters (leading, trailing, or both). Specify one or more characters for **Characters To Trim**. Then select **Trim Leading Characters**, **Trim Trailing Characters**, or both.

Uppercase String

Available for string fields. This function converts the string to uppercase.

Round Number

Available for numeric fields. This function rounds the number as specified. Also specify **Scale**, the number of decimal places.

7.5 Adding a Validation Activity

A validation activity essentially defines a validation report; the report will flag any rows that failed the validation rules that you specify. For each validation rule, you specify whether to halt the recipe or to simply issue a warning. In the case of any validation error, the system generates a workflow task for the assigned role (specified when the recipe is [scheduled](#)).

To add a validation activity to a recipe:

1. Display the recipe.
2. Click **Add Validation Activity**.
3. For **Name**, enter a descriptive name to display within the recipe.
4. For **Target Source**, select the applicable data asset by choosing the schema that describes the asset. The list includes all schemas used in *published* [staging activities](#) within the same recipe.

The system then displays the fields in the staging table, with options to specify validation rules for any of the fields. Note that any field can have at most one validation rule, within a given validation activity.

5. For each field where you want to apply a validation rule, select the field name.

The system then displays options on the right side of the page. Make edits described below and then click **Apply**.

- **Function Type**—See [Validation Function Details](#).
- **Validation Name**—Specify a short name for this field transformation.
- **Result Field Name**—Specify a name for the report field that will store either 1 (if the field value passes the validation rule) or 0 (if it does not). It is useful for the name to indicate the field being examined, for example `CheckItemCount`.
- **Validation Failure**—Specify how the system should respond when it encounters a field value that fails this specific rule. A **Fatal** failure halts the recipe. A **Warning** failure logs a warning. Both kinds of failures are visible in the [Workflow Inbox](#).
- Other options—See [Validation Function Details](#).

6. When you are done editing this activity, click **Save Draft**. Or click **Publish Activity** so that the recipe will use it.
7. Click the **Recipe** link in the upper left to return to the recipe.

7.6 Validation Functions

This section provides reference information on the [validation](#) functions.

Code Table

Available for string and numeric fields. For this function, the field value fails validation if it does not match a value in a code table. Also specify:

- **Code Table Name**—Specify the name of the code table to use.
- **Code Table Key Column**—Specify the name of the key column in this table.

Custom SQL Expression

Available for all types of fields. For this function, the field value fails validation if an SQL predicate expression returns false. (You can, for example, compare the value in this field to the value in another field in the same record.) Also specify:

- **SQL Operator**—Select a comparison operator.
- **SQL Expression**—Enter a scalar SQL expression such as a field name or a constant.

Not Null

Available for all types of fields. For this function, the field value fails validation if it is null.

Numeric Change Within Range

Available for numeric, date, and time fields. For this function, the field value fails validation if the value obtained in this recipe run is too different from the value in the previous run of the recipe. Also specify **Tolerance**—the amount by which the value is permitted to be different from the previous value. This can be either an absolute value or a percentage. (To specify a percentage, include a percent sign at the end.)

If the field is a date field, the unit is a day. If the field is a datetime field, the unit is a second.

Fixed Numeric Range

Available for numeric fields. For this function, the field value fails validation if it falls outside of a fixed range of numeric values. For the **Fixed Numeric Range** function type, also specify:

- **Minimum**— Minimum permitted numeric value.
- **Maximum**— Maximum permitted numeric value.
- **Tolerance**—Specify the tolerance, or the amount by which the value is permitted to be outside of the specified range. This can be an absolute value or a percentage. (To specify a percentage, include a percent sign at the end.)

String Change Within Code Table

Available for string fields. This function is intended for the scenario where there is a code table that represents an ordered sequence such as a set of ratings, and the goal is to check that the field value has not changed too much within that sequence, since the previous successful load of data. For example, in this code table, the rating AAA may be represented as 10, the rating AA+ may be represented as 9, and so on. To use this function type, also specify:

- **Tolerance**—Specify the tolerance, or the amount by which the value is permitted to be outside of the specified range. This can be an absolute value or a percentage. (To specify a percentage, include a percent sign at the end.)
- **Code Table Name**—Specify the name of the code table to use.
- **Code Table Key Column**—Specify the name of the key column in this table. The key values are treated as strings.
- **Code Table Value Column**—Specify the name of the value column in this table. The values in this column must be numeric.

This function examines the previous and current values of the field, both of which are strings such as AAA and AA+, converts them to the corresponding numbers, and compares the numbers, taking the tolerance into account. If the numbers match or are within the tolerance amount of each other, the new field value passes validation.

Value Changed Since Previous

Available for all types of fields. For this function, the field value fails validation if the value is different from the value obtained in the previous successful run of the recipe. Use this rule if you expect the value to always be the same.

7.7 Adding a Reconciliation Activity

A reconciliation activity uses comparison data, which can be either data that you have loaded (and if needed transformed and validated) within the same recipe or data from another data source.

To add a reconciliation activity to a recipe:

1. Display the recipe.
2. Click **Add Reconciliation Activity**.
3. For **Reconciliation Activity Name**, enter a descriptive name to display within the recipe.
4. In the **Primary Source** section, specify the table that you ultimately want to promote and use downstream. To do so:

- a. For **Data Source**, select the location of the table that you want to use. Typically this is **Staging Tables**, because the most common use case is to reconcile data you have loaded within a recipe.
 - b. For **Tables**, select the table.
5. In the **Secondary Source** section, specify the comparison table. To do so:
 - a. For **Data Source**, select the location of the table that you want to use.
 - b. For **Tables**, select the table.
 6. Click **Create Activity**.
 7. On the **Key Map Definition** tab, optionally click the **Report Missing Records from Primary Source (Full Outer Join)** check box. If you select this check box, the comparison will report whenever there are records in the primary source that are not found in the secondary source.

Below this check box, the page initially displays a list of the fields in the primary source, with no information about how these are to be matched to fields in the secondary source.

8. The next step is to specify how to match a row in the primary source to a row in the secondary source. Typically, each source has a primary key (consisting of one or more fields), and the primary keys can be matched to each other; that is a record that has a specific primary key in the one source will be matched to the record that has the same primary key in the second source.

If the sources both have a primary key, the easiest approach is to click **Auto Match Primary Keys**. When you do so, the **Key Map** area shows the primary key fields from the primary source on the left, with the primary key fields from the secondary (comparison) source on the right.

9. Now add fields to the comparison, as follows:
 - a. Click a field name under **Primary Source**—this adds the field to the left column in **Key Map** area.
 - b. Look for the field under **Secondary Source** that should be used for comparison for this field. Click that field name. The system adds this field to the right column in **Key Map** area.

Or click **Try to Match Selected Column**. In this case, the system looks for an appropriately named field in the secondary source.

To remove an item in the **Key Map** area, click its X button.




10. When you are done adding fields, click **Create**. The system then creates the key map for this reconciliation activity. If needed, you can return later and make edits.
11. Click **Reconciliation**.

Now the page displays a table listing the fields from the two sources. Here you specify rules used to compare the values.

12. Briefly review the table for familiarity. The **Primary Field** column lists the fields from the primary source, and the **Secondary Field** column list the fields from the secondary source. This information is the same as in the key map. In addition, note that **Data Type** indicates the data types of the fields from the primary source (which is helpful in choosing a comparison function).
13. If many or most of the fields have the same name in the two tables, click **Quick Match All** to quickly create a set of reconciliation rules that you can then review and edit. With this option, the system automatically creates a rule for each pair of fields that have the same name (including case). Each automatically generated rule includes a unique rule name, a comparison function, a generated field name (to hold the comparison result), and (for numeric comparisons), an initial setting for **Tolerance**.

Some fields may not yet have rules (which is OK).

14. Make edits (as needed) for each rule:

- a. In **Function Type**, choose a comparison function.
 - b. For **Rule Name**, click the Edit  icon and then type a unique rule name.
 - c. For **Result Field**, click the Edit  icon and then type a unique field name. The reconciliation activity will write a 1 or a 0 to this field, to indicate whether reconciliation was successful.
 - d. For **Tolerance** (required for numeric comparison), click the Edit  icon and then type the numeric tolerance.
If the tolerance ends with a percent sign (%), then the system treats it as a percentage.
 - e. If failing this reconciliation check should be a fatal error (halting the recipe), select the **Fatal** check box.
15. When you are done editing this activity, click **Save Draft**. Or click **Publish Activity** so that the recipe will use it.
 16. Click the **Recipe** link in the upper left to return to the recipe.

7.8 Viewing the Staging Table Details

Before you can add a promotion activity, it is helpful to review the structure of the [staging table or tables](#) used by the recipe. In particular, make a note of the field names that you need to use. To see the definition of a staging table:

1. Display the recipe.
2. At the bottom of the page, see the list in **Indices and Staging Table Details**.
This section shows one row for each data source used in the recipe.
3. Click the row for a given data source. The page then lists the generated staging tables associated with this data source.
4. Click a table name in this list. The right side of the page then provides information about the indices on the table, along with details about the fields in this table.
5. Optionally add indices as needed. To add an index, click **New Index** and then specify an index name and select one or more properties to be indexed.

7.9 Adding a Promotion Activity

A promotion activity copies data from the staging table to a final table or final tables, which can be within Data Fabric Studio or which can be in an external database. The activity consists of one or more promotion items, each of which uses custom SQL to delete, insert, or update records of an existing table.

Because the staging table name is generated, it is necessary to refer to the staging table name in an indirect way. Similarly, it is necessary to refer indirectly to the current run of the recipe. Accordingly, the product provides options that insert special tokens that you can use within your SQL; these tokens are automatically replaced when the recipe is run.

To add a promotion activity to a recipe:

1. Display the recipe.
2. Click **Add Promotion Activity**.
3. For **Name**, enter a descriptive name to display within the recipe.

4. For **Run Order**, select or type an integer, which controls the order in which the promotion activities are run, with respect to other promotion activities in the same recipe.
5. For **Promotion Type**, select one of the following:
 - **Internal** — Select this to load data into a table within this database (that is, the database native to Data Fabric Studio).
 - **External** — Select this to load data into a table in another database.

If you select **External**, choose a database for **Target Database**.
6. Click **Submit**.
7. Click **New Item**.
8. Specify the following values:
 - **Run Order**—Specify the order in which this item should be executed, relative to other items in the same promotion activity. Select an integer.
 - **Description**—Enter an optional description of this item, as an aid to yourself and others reviewing this recipe later.
 - **SQL Expression**—Enter an SQL UPDATE, INSERT, or DELETE statement that affects the final table.

The right side of the page provides samples to assist you in writing the query. When you are getting started, it is a good idea to review these and find an example that matches your scenario, to use as a starting place.

Below **SQL Expression**, the page provides options that you can use to insert helpful tokens into the query. To insert a reference to a staging table:

- a. Place the cursor within **SQL Expression** at the position where you need the staging table reference.
- b. For **Placeholders**, select **Staging Activity**.
- c. For **Staging Activities**, select the short name of the staging activity to which the desired staging table belongs. (For example, `mystage`.)
- d. For **Data Source Items**, select the data schema to which the desired staging table belongs. (For example, `samplecsv`.)
- e. Click **Add to SQL**.

This inserts the following string into your query:

```
{mystage}.samplecsv
```

This is a generic reference to the staging table for the `samplecsv`, generated and populated by the `mystage` staging activity within this recipe. When the recipe is run, this token is replaced with the actual name of this staging table.

Similarly, when selecting records from the staging table (to insert or update the final table), you typically select only the records from the current run of the recipe. To modify your query to do this:

- a. If there is no **WHERE** clause yet, move the cursor to the end of **SQL Expression** and type `WHERE` followed by a space.
- b. For **Placeholders**, select **BatchID**.
- c. Click **Add to SQL**.

This inserts the following string into your query:

```
%BatchId={%BatchId}
```

Note that `%BatchId` is the name of the field that stores the recipe run ID, and the token `{%BatchId}` is converted automatically into the ID of the current recipe run.

You can also type these syntaxes directly into the query.

9. Click **Save Activity Item**.
10. If necessary, add other items to this promotion activity.
11. When you are done editing this activity, click **Save Draft**. Or click **Publish Activity** so that the recipe will use it.
12. Click the **Recipe** link in the upper left to return to the recipe.

7.10 Adding a Custom Activity

A recipe can include custom activities. For example, it may be necessary to have a custom activity that retrieves the data from a data source (for example, to place files within the file system). In such cases, two items are needed:

- Custom code in the form of a business host added within the interoperability engine that system uses internally. The business host must be a business operation or a business process that accepts as input a message of type `SDS.DataLoader.CustomActivityRequest`. Details are currently beyond the scope of the documentation.
- Configuration of the recipe to use that code at the applicable phase. To configure a recipe to use custom code:
 1. Display the recipe.
 2. Click **Add Custom Activity**. This displays a dialog box where you specify the details.
 3. For **Name**, enter the name of this activity.
 4. For **Target Business Host**, enter the configuration name of the business host.
 5. For **When to Run**, select **Before Staging**, **Before Transformation**, **Before Validation**, **Before Reconciliation**, **Before Promotion**, or **After Promotion**.
 6. For **Enabled**, select this check box if the activity should be enabled.
 7. Click **Submit**.

7.11 Step Mode

You can run a recipe in step mode, a mode that lets you use a subset of data—perhaps a very small subset—so that you can test the recipe quickly from end to end. In this mode, you specify the number of rows of data to load into the staging tables (and then to use in the subsequent processing).

To configure a recipe for step mode:

1. Display the recipe.
2. Click the **Step Mode** toggle.
3. In the box next to **Step Mode**, select or type the number of records to use.

A recipe remains in step mode until you change the **Step Mode** toggle back to off.

7.12 See Also

- [Introduction to Recipes and Staging Tables](#)
- [Editing and Managing Recipes](#)
- [Managing Recipe Groups](#)
- [Scheduling and Running Tasks](#)
- [Viewing Run History](#) (includes information on viewing dropped records, validation reports, and reconciliation reports)

8

Editing and Managing Recipes

In InterSystems® Data Fabric Studio™, a recipe describes how to load data into the system. This page describes how to edit and manage recipes; another page provides information on [defining](#) them.

For convenience, recipes can be in [groups](#).


8.1 Recipe Changes

One of the key objectives is to permit schemas to change, while preserving your data. [Schema evolution](#) is described in detail elsewhere, but note the following:

1. When you [edit](#) or [reimport](#) a schema, you are implicitly creating a draft that is not yet used by anything. The existing schema is unchanged; existing data is also unchanged.
2. When you [republish](#) a schema, *if* the change affects the structure needed to contain the data, the system automatically increments the version number of the schema and automatically generates a new draft of any recipe staging activity that is affected by this change. The new staging activity draft is not automatically published, and the previous staging activity is unchanged.
3. When you republish any draft staging activities, the system generates a new staging table, considering the structural change that has occurred. This leaves the old table (and its data) unchanged.

Independent from any schema changes, a recipe can also require fields to be added to the staging table. The system handles this automatically.

8.2 Viewing the Recipes

To see all the available recipes, click the Recipes  icon in the application menu. The page displays a list of the recipes; some or all of them might be in [groups](#). In this display, the groups are listed alphabetically, followed by any recipes that are in no groups. Within a group, the recipes are listed alphabetically. You can [specify the group](#), if any, to which a recipe belongs.

Because you might have a large number of recipes, this page provides filter options that can help you find a recipe more quickly.


To display only recipes whose names match a given string, type that string into the Search box. Note that this hides recipes whose names do not match the string, but does not hide any groups.

To filter the recipes so that you see only the recipes with alerts or notifications, click the Filter icon and then click **Only Show Items With Alerts/Notifications**. Then click **Apply**.

This page does not show deleted groups by default. To display deleted groups, click the Filter icon and then click **Show Deleted Groups**. Then click **Apply**.

8.3 Editing a Recipe

To edit a recipe (other than [renaming](#) it):

1. Click the Recipes  icon in the application menu.
2. If the recipe is in a group, expand that group.
3. Click the recipe. The system then displays the current definition, including any draft activities.

If there is a draft activity, you can choose to display the published version or the draft version (which is shown as a separate item as in the following example).



Now you can make any of the following edits:

- To add a new activity, click **Add Staging Activity**, **Add Transformation Activity**, **Add Validation Activity**, **Add Reconciliation Activity**, or **Add Data Promotion Activity**, and then continue as described in [Defining Recipes](#).

The new activity will not take effect until you publish it via the **Publish Activity** button.


- To edit an existing activity, click it. Notes:
 - If you click a draft that the system generated because of [schema evolution](#), the system displays a message that explains why it generated this new draft.
 - If you are viewing a published activity, you must click **Edit Activity** to make it editable.
 - If you are viewing a published activity *and* there is a current draft for that activity, the **Edit Activity** button displays the draft in edit mode.

Make changes as needed and save them via the **Save Draft** button.

The changes will not take effect until you publish it via the **Publish Activity** button.

8.4 Publishing an Activity


To publish (or republish) an activity within a recipe:

1. Click the Recipes  icon in the application menu.
2. If the recipe is in a group, expand that group.

3. Click the recipe. The system then displays the current definition, including any draft activities.
If there is a draft activity, you can choose to display the published version or the draft version.
4. Click the activity that you want to republish.
5. Click **Publish Activity**.

8.5 Disabling an Activity

To disable an activity within a recipe (so that this activity is skipped if the recipe is used):

1. Click the Recipes  icon in the application menu.
2. If the recipe is in a group, expand that group.
3. Click the recipe. The system then displays the current definition, including any draft activities.
If there is a draft activity, you can choose to display the published version or the draft version.
4. Click the activity that you want to disable.
5. Click the slider in the **Actions** column for that activity. Then click **Disable Current Activity**. Here's an example (before):




After:




8.6 Enabling an Activity

To enable an activity that has previously been disabled:

1. Click the Recipes  icon in the application menu.
2. If the recipe is in a group, expand that group.
3. Click the recipe. The system then displays the current definition, including any draft activities.
If there is a draft activity, you can choose to display the published version or the draft version.
4. Click the activity that you want to enable.
5. Click the slider in the **Actions** column for that activity.

8.7 Deleting an Activity

To delete an activity in a recipe.



1. Click the Recipes  icon in the application menu.
2. If the recipe is in a group, expand that group.
3. Click the recipe. The system then displays the current definition, including any draft activities.
4. Click the activity that you want to delete.
5. Click **Delete**. (Or if you are deleting a draft activity, click **Delete Draft**.)

Note that system-generated drafts cannot be deleted; system-generated drafts are created when, for example, an associated schema has changed.

6. Click **Delete** again.



8.8 Renaming a Recipe

To rename a recipe:

1. Click the Recipes  icon in the application menu.
2. If the recipe is in a group, expand that group.
3. Click the Menu  icon next to the recipe name.
4. Click **Rename**.
5. Enter new values as needed for **Recipe Name** and **Short Name**.
6. Click **Rename** to save the changes.

8.9 Moving a Recipe Into or Out of a Group



You can move a recipe into a group or out of all groups via drag and drop. The system also provides a menu option, as follows:

1. Click the Recipes  icon in the application menu.
2. If the recipe is currently in a group, expand that group.
3. Click the Menu  icon next to the recipe name.
4. Click **Move to other group**.
5. Select a group name (or select **Ungrouped** if this recipe should not be in any group).

Note that you cannot move a recipe into a group that has been deleted.



8.10 Copying a Recipe

To copy a recipe:

1. Click the Recipes  icon in the application menu.
2. If the recipe is in a group, expand that group.
3. Click the Menu  icon next to the recipe name.
4. Click **Copy**.
5. Enter new values as needed for **Recipe Name** and **Short Name**.
6. Optionally select **Continue Incremental Loads**. This option is useful if the new recipe is intended to carry on the same work as the one it was copied from. If this option is selected and if the original recipe loaded data incrementally and its last load was successful, that data will be checked and the new recipe will pick up where the original left off.
7. Click **Copy Recipe** to save the changes.

8.11 Disabling a Recipe



To disable a recipe, so that it cannot be run:

1. Click the Recipes  icon in the application menu.
2. If the recipe is in a group, expand that group.
3. Click the Menu  icon next to the recipe name.
4. Click **Disable**.
5. Click **Disable** to confirm.

When you disable a recipe, the system disables any scheduled tasks that run the recipe.



8.12 Enabling a Recipe

To enable a recipe that has previously been disabled:

1. Click the Recipes  icon in the application menu.
2. If the recipe is in a group, expand that group.
3. Click the Menu  icon next to the recipe name.
4. Click **Enable**.

8.13 Purging Old Recipe Data

To purge old data for a recipe:

1. Click the Recipes  icon in the application menu.
2. If the recipe is in a group, expand that group.
3. Click the Menu  icon next to the recipe name.
4. Click **Reset**.
5. Click one of the following options:
 - **Full Reset** deletes all staging tables (both old and current) and deletes the recipe run history.
 - **Reset All But Current** just deletes the old staging tables, leaving the current staging tables as is. The old staging tables correspond to previous versions of the recipe.

Neither option affects the *definitions* of the current staging tables.



When you click an option, the right side of the dialog box provides additional detail on what will be deleted.

6. Click **Reset Recipe** to confirm.

CAUTION: This step is irreversible.

8.14 Deleting a Recipe

To delete a recipe:

1. Click the Recipes  icon in the application menu.
2. If the recipe is in a group, expand that group.
3. Click the Menu  icon next to the recipe name.
4. Click **Delete**.
5. Click **Delete** to confirm.

When you delete a recipe, the system disables any scheduled tasks that run the recipe.

8.15 Recovering a Recipe

To recover a recipe that was previously deleted:



1. Click the Menu  icon next to the recipe name.

2. Click **Recover**.

The recipe is immediately restored and now can be used.

8.16 Permanently Deleting a Recipe

When you delete a recipe, it is possible to recover it. You can also permanently delete a recipe (so that it cannot be recovered). To do so:

1. [Delete the recipe](#) as previously described or [delete the group](#) to which it belongs.
2. If you deleted the group:
 - a. Click the Filter  icon.
 - b. Then click **Show Deleted Groups** and then click **Apply**.
3. Click the Menu  icon next to the recipe name.
4. Click **Permanently Delete**.
5. Click **Delete** to confirm.

When you permanently delete a recipe, the system also deletes the following items:

- Any scheduled tasks that run the recipe.
- All run history associated with the recipe.
- The staging tables that were generated for the recipe.

8.17 See Also

- [Defining Recipes](#)
- [Managing Recipe Groups](#)
- [Scheduling and Running Tasks](#)
- [Handling Task Errors](#)
- [Viewing Run History](#)


9

Managing Recipe Groups

Ultimately there may be a large number of [recipes](#), so the product provides an option to group them. When you display the recipes, the groups are listed alphabetically, followed by any recipes that are in no groups. Within a group, the recipes are listed alphabetically.



9.1 Creating a Recipe Group

To create a recipe group:

1. Click the Recipes  icon in the application menu.
2. Click **New group**.
3. Type a unique group name and press **Return** (or click the check mark).

9.2 Renaming a Recipe Group

To rename a recipe group:



1. Click the Recipes  icon in the application menu.
2. Click the Menu  icon next to the recipe group name.
3. Click **Rename**.
4. Type the new name and click the check mark.

The new name must be unique.

9.3 Deleting a Recipe Group




Tip: Before deleting a recipe group, identify any recipes in the group that you want to keep, and move those out of this group. When you delete a recipe group, that action deletes the recipes in it. They can be recovered after that step, but the easiest procedure is to move them first.

To delete a recipe group:

1. Click the Recipes  icon in the application menu.
2. Click the Menu  icon next to the recipe group name.
3. Click **Delete**.
4. Click **Delete Group** to confirm this action.

9.4 Recovering a Deleted Recipe Group

To recover a recipe group that was previously deleted:

1. Click the Recipes  icon in the application menu.
2. Click the Filter  icon.
3. Click **Show Deleted Groups** and then click **Apply**.
This action displays the deleted groups.
4. Click the Menu  icon next to the recipe group name.
5. Click **Recover**.
6. Click one of the following options, depending on whether you also want to recover the recipes in this group:
 - **Recover the group only**
 - **Recover the group and all of the group's recipes**

Then click **Submit**.

Note that you can also [recover individual deleted recipes](#).

9.5 See Also

- [Defining Recipes](#)
- [Editing and Managing Recipes](#)

10

Defining Snapshots

Snapshots are static copies of data in InterSystems® Data Fabric Studio™, for future review. This page describes how to define them; [another page](#) describes how to edit and manage them.

Once you have defined snapshots, you can [run them](#) (loading data into the system), either on a schedule or manually.

10.1 Uses of Snapshots

With snapshots, you can easily save data for later inspection by regulators; the snapshot can pull data from multiple tables as needed, and the system applies a tag to the records.

When you have run a snapshot multiple times, applying a different tag each time, you can examine how that data changes over time. In particular, you can build a cube on the snapshot data, using the tag values as a dimension.

10.2 Snapshot Tables

For each snapshot, the system will generate a table. The table name is based on the snapshot definition, and the columns are based on the query used in the snapshot definition. The table name format is as follows:


`halp_snapshot_snapshot.tablenameinteger`

Where *snapshot* is the **Short Name** of the snapshot and *tablename* is the short table name, both of which you provide when defining the snapshot. At the end of the name, *integer* indicates the version number. (Note that `halp_snapshot` is a default part of this name but is [configurable](#).)

If you change a snapshot definition so that it defines a different set of columns, the system automatically increments the version number and generates a new table that has the required columns. For example, when you define a snapshot, the generated table name may be `halp_snapshot_Sample.SampleTable1`, but if you redefine the query, the new generated table name becomes `halp_snapshot_Sample.SampleTable2`, and so on.

10.3 Defining a Snapshot

To define a snapshot:

1. Click the Snapshot  icon in the application menu.
2. Click **New Snapshot**.

The system displays a dialog box where you specify initial information.
3. Specify the following information:
 - **Name**—*Required*. Specify a unique name for this snapshot.
 - **Short Name**—*Required*. Specify a unique short name for this snapshot, to be used within the fully qualified name of the generated table.
 - **Table Name**—*Required*. Specify a unique short table name (also to be used within the fully qualified name of the generated table).
 - **Tag**—*Required*. Specify a short string that becomes a tag applied to the snapshot when the snapshot is executed. As an example, a tag could indicate the state of the data, such as Preliminary or Final.
 - **Description**—Type a description of this snapshot and its purpose.

Except for **Short Name** and **Table Name**, you can edit these values later as well.

4. Click **Submit**.

Now you can define the rest of the snapshot.
5. To specify the query that returns the data for the snapshot:
 - a. For **SQL Statement**, specify an SQL SELECT statement that retrieves the values you want in the snapshot. The query can refer to multiple tables. The query can use * where that is syntactically valid.
 - b. Click **Parse SQL Statement**.

The system then tries to parse the SQL statement and determine the structure of the table it will generate and populate. If the system can parse the SQL, the **Columns** section then displays the fields of the new table, and the **Indices** section displays the indices it will generate for the new table.

If the system cannot parse the SQL, a warning is displayed, and you can edit the query.

6. Optionally add custom indices to this table. To add an index:
 - a. Click **New Index**.
 - b. For **Index Name**, type a unique index name.
 - c. For **Property 1**, select the field to index.

To create an index on multiple fields, repeat with **Property 2** and so on.
7. If there is a table that you want to lock while performing this snapshot (in addition to the table or tables used in **SQL Statement**):
 - a. Click **Add Table**.
 - b. For **Table Name**, type the fully qualified table name (not just the short name).
 - c. Click **Submit**.
8. Click **Save**.

Notice that after you have saved a snapshot for the first time, the **Tables To Lock For Snapshot** section lists the table or tables to which **SQL Statement** refers.

Tip: After you run a newly defined snapshot, use the [SQL Explorer](#) to verify that the snapshot definition has captured the data you need to save.

10.4 See Also

- [Using the SQL Explorer](#)
- [Editing and Managing Snapshots](#)
- [Scheduling and Running Tasks](#)
- [Handling Task Errors](#)
- [Viewing Run History](#)

11

Editing and Managing Snapshots

Snapshots are static copies of data in InterSystems® Data Fabric Studio™, for future review. Once you have defined snapshots, you can [run them](#), either on a schedule or manually. This page describes how to edit and manage them; [another page](#) describes the basics of defining them.

11.1 Viewing the Snapshots Dashboard

To view the Snapshots Dashboard, click the Snapshot  icon in the application menu.

This page enables you to see all the snapshots. For each snapshot, the page displays the following information.

- **Status**—Status of the snapshot, which can be **Idle**, **Running**, or **Error**.
- **Snapshot Name**—Name of the snapshot.
- **Snapshot Description**—Description of the snapshot.
- **Table Name**—Name of the table in which the current version of this snapshot is stored.
- **Life Cycle**—Displays the life cycle state of the snapshot, which can be **Active** (the initial state), **Disabled**, or **Deleted**.


Only **Active** snapshots can be [scheduled](#).

Click any snapshot in order to view it or make changes.

To filter the display, use the options above the column headers.

11.2 Editing a Snapshot

To edit a snapshot:

1. Click the Snapshot  icon in the application menu.
2. Click the snapshot you want to edit.
The snapshot definition is displayed.
3. Click **Edit**.

4. Make changes in the same way as when you [create](#) a snapshot.



If you change the SQL query used in the snapshot, click **Parse Query** so that the system can update the structure of the associated [table](#), if needed. If the snapshot definition now defines a different set of columns, the system automatically generates a new table that has the required columns. The older table is left alone. The system displays an informational message about this change.

5. Click **Save**.

Or click **Delete Draft** to cancel your edits.

11.3 Retagging a Snapshot Run


If you have [run](#) a snapshot, you can change that tag that is associated with any particular run of that snapshot. To do so:

1. Click the Snapshot  icon in the application menu.
2. Click the snapshot.
The snapshot definition is displayed.
3. Click **Snapshot History**.
4. In the row for the run that you want to retag, click the Edit  icon in the **Tag Name** column.
5. Type a new tag and press **Enter**.

Also see [Viewing the Run History of a Snapshot](#).

11.4 Disabling a Snapshot

To disable a snapshot so that it cannot be run:

1. Click the Snapshot  icon in the application menu.
2. Click the snapshot you want to disable.
The snapshot definition is displayed.
3. Click **Disable**.
4. Click **Disable** to confirm.

11.5 Enabling a Snapshot


To enable a previously disabled snapshot:

1. Click the Snapshot  icon in the application menu.

2. Click the snapshot you want to disable.
The snapshot definition is displayed.
3. Click **Enable**.

11.6 Deleting a Snapshot

To delete a snapshot:

1. Click the Snapshot  icon in the application menu.
2. Click the snapshot you want to delete.
The snapshot definition is displayed.
3. Click **Delete**.
4. Click **Delete** to confirm.

11.7 See Also

- [Defining Snapshots](#)
- [Scheduling and Running Tasks](#)
- [Viewing Run History](#)


12

Managing Table Indices

This page explains how to add custom indices to tables in InterSystems® Data Fabric Studio™ and explains how the system automatically builds indices as needed.

12.1 Adding an Index to a Staging Table

Within the definition of a recipe, you can add custom indices to the generated staging table. To do so:

1. Click the Recipes  icon in the application menu.
2. If the recipe is in a group, expand that group.
3. Click the recipe name.
4. At the bottom of the page, see the list in **Indices and Staging Table Details**.
This section shows one row for each data source used in the recipe.
5. Click the row for a given data source. The page then lists the generated staging tables associated with this data source.
6. Click a table name in this list. The right side of the page then provides information about the indices on the table, along with details about the fields in this table.
7. For **Index Name**, type a unique index name.
8. For **Property 1**, select the field to index.


To create an index on multiple fields, repeat with **Property 2** and so on.

9. Because there is now a new index, the **Build Indices** button is enabled. You can click that to immediately start building this index for any existing data in this staging table.

However, even if you do not click **Build Indices**, the system automatically schedules a background task that will build this index for any existing data in this staging table. The system will also automatically index new data.

12.2 Adding an Index to a Snapshot

Within the definition of a snapshot, you can add custom indices to the generated snapshot table. To do so:

1. Click the Snapshot  icon in the application menu.
2. Click the snapshot you want to edit.
3. Click **Edit**.
4. Click **New Index**.
5. For **Index Name**, type a unique index name.
6. For **Property 1**, select the field to index.

To create an index on multiple fields, repeat with **Property 2** and so on.

When you add an index, the system automatically detects this change and schedules a background task that will build this index for any existing data in this table. The system will also automatically index new data.

12.3 Index Build Pending Message

Sometimes the process of building an index can be time-consuming and it is necessary to wait before performing recipe runs or snapshots that use the table being indexed. In such a case, the system displays a message indicating that it is in the process of building a required index. The system prevents you from prematurely using the data.

12.4 See Also

- [Editing and Managing Snapshots](#)
- [Editing and Managing Recipes](#)

13

Defining Entities and Calendars


A key part of InterSystems® Data Fabric Studio™ is the ability to [schedule](#) automated processing, and these schedules generally need to align with relevant calendars. For example, a report may need to get run on every business day but not on holidays or weekends. However, not all regions and not all organizations necessarily use the same calendar.

To simplify the handling of calendars, the system provides the concept of entities. An *entity* defines a calendar; specifically, each entity has a time zone, a fiscal calendar, a business week, and a holiday schedule. An entity can be a child of another entity and can inherit some or all of that data from its parent.

In practice, an entity can represent a region, a subsidiary, or an organization. In any case, the entity defines a set of calendar information that you want to rely on, for scheduling purposes.

13.1 Defining an Entity


To create a top-level entity:

1. Click the Entity Master  icon in the application menu.
2. Click **Add Entity**.
The system displays a dialog box where you can specify the details.
3. Enter the following information:
 - **Enable Entity**—Select this if you want the entity to be available for use in [scheduling](#).
 - **Entity Description**—Enter the name of this entity, which must be unique.
 - **Time Zone**—Select the time zone to which this entity belongs.
 - **Fiscal Year Start Month**—Select the first month of the fiscal year for this entity.
 - **Fiscal Year End Month**—Select the last month of the fiscal year for this entity.
 - **Business Week First Day**—Select the first day of the business week for this entity.
 - **Business Week Last Day**—Select the last day of the business week for this entity.
4. Click **Submit**.

Now you can [define the holiday calendar](#) for this entity. You can also [define child entities](#) that use parts of this entity's calendar data.

13.2 Defining a Child Entity


To create a child entity:

1. Click the Entity Master  icon in the application menu.
2. Select the entity to which you want to add a child.
3. Click **Add Child Entity**.

The system displays a dialog box where you can specify the details.
4. Enter the following information:
 - **Enable Entity**—Select this if you want the entity to be available for use in [scheduling](#).
If the parent entity is not enabled, this entity cannot be enabled.
 - **Entity Description**—Enter the name of this entity, which must be unique within the parent entity.
 - **Use Parent Time Zone?**—Select this if you want the entity to use the time zone defined by the immediate parent entity.
 - **Time Zone**—Select the time zone to which this entity belongs. (Not available if **Use Parent Time Zone** is selected.)
 - **Use Parent Fiscal Calendar?**—Select this if you want the entity to use the fiscal year start and end month defined by the immediate parent entity.
 - **Fiscal Year Start Month**—Select the first month of the fiscal year for this entity. (Not available if **Use Parent Fiscal Calendar** is selected.)
 - **Fiscal Year End Month**—Select the last month of the fiscal year for this entity. (Not available if **Use Parent Fiscal Calendar** is selected.)
 - **Use Parent Business Week?**—Select this if you want the entity to use the business week start and end days defined by the immediate parent entity.
 - **Business Week First Day**—Select the first day of the business week for this entity. (Not available if **Use Parent Business Week?** is selected.)
 - **Business Week Last Day**—Select the last day of the business week for this entity. (Not available if **Use Parent Business Week?** is selected.)
 - **Use Parent Holiday Calendar?**—Select this if you want the entity to use the holiday calendar of the immediate parent entity. Note that you can [override the calendar details](#).
5. Click **Submit**.

13.3 Editing an Entity

To edit an entity:

1. Click the Entity Master  icon in the application menu.
2. Select the entity you want to edit.


The entity definition is displayed, along with its [holiday calendar](#).

3. Make changes as needed, modifying the values previously described for [top-level entities](#) and [child entities](#).
4. Click **Save**.

Note: Changes to an entity do not affect any tasks that are currently running, because the entity for a task is checked at the start of the task run.

13.4 Editing a Holiday Calendar

To edit the holiday calendar for an entity:

1. Click the Entity Master  icon in the application menu.
2. Select the entity (first expanding the entity tree if necessary).

The system then shows the applicable holiday calendar, next to the entity details. If the selected entity is a child entity that inherits calendar information from its parent, the **Status** column displays **Inherited** for each inherited holiday.

Here you can do the following:

- Filter the display of holidays. To do so, select a year from the dropdown list above the table of dates. Or type the name (or partial name) of a holiday into the box above the table of dates. In either case, the display is immediately filtered.
- Add an entry. To do so:
 1. Click **Add Holiday**.
 2. For **Holiday Description**, enter a short holiday name.
 3. For **Holiday Date**, select a date.
 4. Click **Submit**.
- Remove an entry. To do so:
 1. Click **Delete** in the row for that entry.
 2. Click **Remove Holiday** to confirm.

You cannot remove an inherited entry, but you can deactivate it; see the next item.


- Deactivate an inherited entry. To do so, click the slider in the first column of row for that entry, moving the slider to the off position. The **Status** column will then display **Deactivated** for this holiday.
- Reactivate a previously deactivated (inherited) entry. To do so, click the slider in the first column of row for that entry, moving the slider to the on position. The **Status** column will then display **Inherited** for this holiday.

Each of these changes is immediately saved as part of the entity definition.

Note: Calendar changes do not affect any tasks that are currently running, because the entity for a task is checked at the start of the task run.

13.5 Deleting an Entity

To delete an entity:

1. Click the Entity Master  icon in the application menu.
2. Select the entity (first expanding the entity tree if necessary).
The entity definition is displayed.
3. Click **Delete**.
4. Click **Delete Entity** to confirm.

13.6 See Also

- [Scheduling and Running Tasks](#)

14


Scheduling and Running Tasks

To schedule recipes and other tasks, you use the Business Scheduler, which you can also use to run tasks manually. There are three types of tasks:

- **Recipe**, which runs a single [recipe](#)
- **AtScaleCube**, which rebuilds a [cube](#)
- **Snapshot**, which performs a [snapshot](#) run

14.1 Scheduling a Task

To schedule a task:

1. Click the Business Scheduler  icon in the application menu.
The system displays a table that lists the currently defined schedule items; if needed, use the [filters](#) at the top of the table to narrow down what is shown.
2. Click **Schedule Task**.
The system displays a table that lists the items that can be scheduled. If needed, use the filters at the top of the table to narrow down what is shown.
3. Click the row corresponding to the item to schedule.
4. Specify the following information:
 - **Task Description**—*Required*. Type a brief description for this task.
 - **Entity**—*Required*. Select the [entity](#) whose calendar information you want to use.
 - **Enabled**—Optionally clear this check box if you want to disable this task (for example, if you are not ready to run or schedule it).
 - **Tag**, **Dependency Expression**, and **Dependency Inactivity Timeout**—See [Managing Task Dependencies](#).
 - **Workflow Role for Handling Exceptions**—*Required*. Select the [role](#) that should receive any [workflow items](#) in case of exceptions related to this task.
 - **Email Distribution List for Error Notifications**—Select the [email distribution list](#) that should receive messages when this task encounters an error.

- **Email Template for Error Notifications**—Select the [email template list](#) to use when sending messages when this task encounters an error. This is required if you select **Email Distribution List for Error Notifications** and is hidden otherwise.
- **Email Distribution List for Success Notifications**—Select the [email distribution list](#) that should receive messages when this task runs successfully.
- **Email Template for Success Notifications**—Select the [email template list](#) to use when sending messages when this task run successfully. This is required if you select **Email Distribution List for Success Notifications** and is hidden otherwise.

Then for **Scheduling Details**, either select the **Manually Run** check box *or* specify the following information:

- **Type of frequency to run at** —*Required.* Select the option that best describes how often to run this task.
Depending on your selection, the page may show additional items to control the scheduling, such as the specific days of the week, the specific months, and so on.
- **On Holidays** —*Required.* Select the option that describes how to handle holidays: **Don't Run**, **Run Anyway**, **Run Next Business Day**, **Run Previous Business Day**.
- **Run on the Following Days of the Week** —*Required if task runs every calendar day or every business day.* Select each day on which the task should be run.
- **Start Time on Scheduled Days**—*Required.* Select the time of day when the task should be started (using the local time zone defined for the associated entity).
- **End Time on Scheduled Days**—*Required if task runs more than once a day.* Select the time of day after which the task should not be started (using the local time zone defined for the associated entity).
- **Schedule Task End Time**—Optionally select this if you want the task to stop running at a specific date and time. If you select this option, also specify **Task End Date** and **Task End Time**.

The system displays additional fields to prompt you for information as needed. For example, if **Type of frequency to run at** is **Run Every X Minutes**, the system displays the field **Run Every X Minutes**, which you use to specify the number of minutes.

If you select **Manually Run**, the task is not run until a user chooses to run it.

5. Click **Save Task**.

14.2 Managing Task Dependencies

There may be specific tasks that you want to run only after other tasks have been run, on the same day; the system provides a way to manage these dependencies. To specify that one task (task A) should not be run until another task (task B) has completed, you use the following system:

- Decide on a unique, short tag that describes this dependency (for example, `backup`). Note that tags are case-sensitive.
- When scheduling the task or tasks that need to occur first (in our example, task B), specify the **Tag** field, using the tag you have decided on.
- When scheduling the dependent task (task A), specify **Dependency Expression**, which uses a special kind of syntax. In its simplest form, the expression is as follows:

```
TODAY ( tag )
```


Where *tag* is a tag. For example:

```
TODAY(backup)
```

This expression means that the task should not be started until the completion (today) of a task that is tagged with the backup tag.



You can combine dependency expression via the keywords AND and OR, along with parentheses. For example:

```
TODAY(tagA) OR (TODAY(tagB) AND TODAY(tagC))
```

- In addition to specifying **Dependency Expression**, you can specify **Dependency Inactivity Timeout** in seconds.



14.3 Modifying a Task

To modify a task definition (for example, to change the schedule details):

1. Click the Business Scheduler  icon in the application menu.
The system displays a table that lists the items that are currently scheduled. If needed, use the filters at the top of the table to narrow down what is shown.
2. Click the three dots  in the row for the task. This displays a menu.
3. Click **Edit Task**.
4. Modify details here in the same way as when you [schedule](#) a task.
5. Click **Save Task**.
The system runs the task (or attempts to) within the next minute. When it does so, it updates the **Run History** section of this page.

14.4 Running a Task Manually



To run a task manually:

1. Click the Business Scheduler  icon in the application menu.
The system displays a table that lists the items that are currently scheduled. If needed, use the filters at the top of the table to narrow down what is shown.
Tasks that can be run manually are displayed here with **Next Run** as **Manual**; these are tasks [defined](#) with the **Manually Run** check box selected.
2. Click the three dots  in the row for the task. This displays a menu.
3. Click **Open Task Details**.
4. Click **Manually Run Now**.
5. Click **Run Task** to confirm.

The system runs the task (or attempts to) within the next minute. When it does so, it updates the **Run History** section of this page.

14.5 Aborting a Task

To stop a task that is currently running:



1. Click the Business Scheduler  icon in the application menu.
2. Click the three dots  in the row for the task. This displays a menu.
3. Click **Open Task Details**.
4. Click **Abort Task**.
5. Click **Abort Task** to confirm.

The system runs the task (or attempts to) within the next minute. When it does so, it updates the **Run History** section of this page.

Also see [Handling Task Errors](#).

14.6 Deleting a Task

To delete a task definition:

1. Click the Business Scheduler  icon in the application menu.
2. Click the three dots  in the row for the task. This displays a menu.
3. Click **Open Task Details**.
4. Click **Delete**.
5. Click **Delete** to confirm.

14.7 See Also

- [Defining Entities and Calendars](#)
- [Defining Recipes](#)
- [Defining Snapshots](#)
- [Defining Cubes](#)
- [Filtering and Customizing the Business Scheduler](#)
- [Handling Task Errors](#)

15

Handling Task Errors (Using the Workflow Inbox)

This page describes what happens and how to proceed when an error occurs during a recipe, snapshot run, or other task.


15.1 When the System Encounters an Error

When the system encounters an error while executing a task (either scheduled or manually run), the system does several things:

- Sends email to a specified email distribution list (as specified in the [task schedule](#)).
- (In most cases) Generates a workflow task, and assigns the task to the applicable workflow role (as specified in the [task schedule](#)).
- Updates the display within Business Scheduler to indicate the error condition.
- Updates the display of the associated recipe or snapshot to indicate the error condition.


15.2 First Steps

The first step is to claim ownership of the problem and determine its cause, as follows:

1. Click the Workflow Inbox  icon.
2. Click **Unassigned**.
The system displays a list of the unassigned tasks.
3. Click the task to display its details.
4. Click **Accept**. This assigns the task to you.
5. Optionally type a note into the **Comment** field and click **Save**.
6. Then see [Addressing a Task](#).

15.3 Viewing the Workflow Inbox

The Workflow Inbox shows all the workflow tasks. To display it:

1. Click the Workflow Inbox  icon in the application menu. The system initially displays all tasks assigned to you, within the searchable task list.
2. Optionally click any of the following options, to display a different set of tasks:
 - **All** displays all tasks, for all roles, showing both incomplete tasks and completed tasks.
 - **My Tasks** displays tasks assigned to any of the roles that you belong to.
 - **Unassigned** displays any unassigned tasks.
 - **Completed** displays all completed tasks.
 - **My Completed** displays all completed tasks assigned to any of the roles that you belong to.
3. Optionally type into the search bar above the task list. The system then displays only the tasks whose text contains a match for your search string.

When you click a task, the system displays details for that task.

15.4 Addressing a Task

To address a workflow task, [first](#) assign the task to yourself. Read the message to determine what the error is and how you plan to correct it. For example, you may need to do any of the following:

- Ensure that the file to be loaded has the correct filename pattern and exists in the correct location (in the case of recipes involving a file-based data source).
- Correct the data being loaded (for example, if a value is out of range or is the wrong type). This may involve making a correction in a source system.
- Reimport a schema, if the new data does not match the schema in the system.
- Adjust a schema (changing data types or required fields, for example).
- Adjust a step in a recipe.
- Adjust the query used by a snapshot.

Then, depending on the kind of underlying activity, you may have some or all of the following options within the workflow task:

- **Abort**—Halts the task without retrying it.
Once you have aborted a task, you cannot retry it.
- **Retry**—Retries the task. This is suitable if the data has been corrected or if a missing file has been made available. It is not useful if you have made changes to a schema or to a recipe.
- **Retry All with New Version**—Retries the task with the new version of the recipe, considering new versions of the failed activity and of any later activities. The system will retry the recipe from the point where it failed, using the updated version of that activity and updated versions of any subsequent activities.

- **Retry Failed Activity with New Version**—Retries the task with the new version of the recipe, considering only the new version of one activity in the recipe. The system assumes that the new activity is the one where the failure occurred, and it will retry the recipe from the point where it failed, using the updated version of that one activity. It uses the original version of any other activities in the recipe.

15.5 Relinquishing a Task

If a task is assigned to you and another person should address the task instead of you, you should relinquish the task so that the other person can accept it. To relinquish a task:

1. Display the task.
2. Click **Relinquish**.

The task immediately becomes unassigned. If the task had a comment, the comment is removed. Now another person can accept the task.

15.6 See Also

- [Scheduling and Running Tasks](#)


16

Viewing Run History


You can view the history and current status for any recipes and snapshots that you have run or are currently running. The recipe run details provide information on dropped records, validation errors, and reconciliation errors.

16.1 Run History of a Recipe

To view the run history of a recipe:

1. Click the Recipes  icon in the application menu.
2. Click the recipe.
3. Click **View Run History**.

This displays the **Recipe Run History** page, which provides summary information for all runs of the selected recipe, as well as a link to the details for each run:

- **Snapshot ID**—Internal identifier of the snapshot.
- **Start Time**—Indicates the date and time when this run of the recipe was started.
- **End Time**—Indicates the date and time when this run of the recipe was completed.
- **Error Details**—Displays error details from this run, if any.
- **Status**—Indicates the [status of the task](#).
- **Visual Trace**—Provides a link to a trace session (for internal use by your solution partner).
- **Record Report**—Provides a  link to the [details](#) for this recipe run.

16.2 Recipe Run Details and Downloadable Reports

To display the details of a recipe run:

1. Click the Recipes  icon in the application menu.


2. Click the recipe.
3. Click **View Run History**.
4. Click the  link in the **Record Report** column for the run of interest.

This displays a page that reports on all activities within the recipe. The page has expandable sections, by activity type, so that you can display only the data of interest to you.

When you view the validation and reconciliation sections, you have the option of downloading the validation and reconciliation reports, respectively.

16.3 Run History of a Snapshot

To view the run history of a snapshot:

1. Click the Snapshot  icon in the application menu.
2. Click the snapshot.
3. Click **Snapshot History**.

This displays the **Snapshot History** page, which presents the following information:

- **Snapshot ID**—Internal identifier of the snapshot.
- **Start Time**—Indicates the date and time when this snapshot was started.
- **End Time**—Indicates the date and time when this snapshot was completed.
- **Duration**—Indicates how long the task took (or has taken).
- **Total Records**—Indicates the number of records in this snapshot.
- **Status**—Indicates the [status of the task](#).
- **Error**—Displays error details from this run, if any.
- **Visual Trace**—Provides a link to a trace session (for internal use by your solution partner).
- **Tag Name**—Displays the tag applied to this task.

You can use the filters above the table column to filter the display.

Here you can [retag a snapshot run](#).

16.4 Run History from the Business Scheduler

You can also view the run history of any task from within the Business Scheduler.

To view the run history of any task:

1. Click the Business Scheduler  icon in the application menu.

The system displays a table that lists the currently defined schedule items; if needed, use the [filters](#) at the top of the table to narrow down what is shown.

2. Click the description of the task.

The bottom of the page includes the **Run History** section, which presents the following information:

- **Status**—Indicates the [status of the task](#).
- **Event Time**—Indicates the date and time when this task was started.
- **Elapsed**—Indicates how long the task took (or has taken).
- **Run by User**—Indicates which user ran this task.
- **Error Details**—Displays error details from the last run, if any.
- **Visual Trace**—Provides a link to a trace session (for internal use).

16.5 Task Status

The status of any task can be any of the following:

- **Idle**—The task is not running.
- **Starting**—The task is starting.
- **RunningWait**—The task is running but is in a waiting state.
- **Running**—The task is actively running.
- **Error**—The task has encountered an error and has stopped. See [Handling Task Errors](#).
- **Complete**—The task has completed.
- **Aborted**—A user [aborted](#) the task (or the system automatically aborted the task).

16.6 See Also

- [Handling Task Errors](#)
- [Scheduling and Running Tasks](#)
- [Purging Old Recipe Data](#)

17

Filtering and Customizing the Business Scheduler

Depending on the number of items defined in the [Business Scheduler](#), you may find it helpful to filter the display, to customize the display, or both.

17.1 Default Display

By default, the Business Scheduler lists all scheduled tasks and displays the following columns of information about those tasks:

- **Status**—Indicates the [status of the task](#).
- **Task Description**—The short name of the task (generally the name of the [recipe](#) or [snapshot](#) that the task is based on).
- **Enabled**—Indicates whether this task is currently enabled.
- **Type**—Indicates the task type: **Recipe**, **Snapshot**, or **AtScaleCube**.
- **Resource Name**
- **Resource Status**—If the task is running, this indicates the current activity of the task. If the task is not running, this indicates the status of the last run of this task.
- **Exceptions**—If the system encountered an error during its most recent task run, this field displays an error message and includes a link to the [Workflow Inbox](#), where you can see the details.

For example:



See [Handling Task Errors](#).

- **Entity**—The [entity](#) whose schedule this task uses.
- **Last Run**—The date and time when this task was last run.
- **Elapsed**—The length of time taken for this task on its last run.
- **Next Run**—Indicates when or how the task will next be run. This is either a future date and time or is simply the word **Manual** (which means that this task is defined to be run manually).

- **Tag**—The [tag](#) applied to this task.
- **Dependencies**—The [dependency expression](#) for this task, if any.

17.2 Filtering the Display


To filter the display, use the dropdown menus and type-in boxes above the table.


To clear all filters, click the grid  icon and then click **Clear Filters**.

17.3 Customizing the Columns

You can customize the columns that the Business Scheduler displays. This customization does not affect other users, and this customization is preserved until you modify it or restore the default set of columns.

To customize the columns:

1. Click the grid  icon and then click **Toggle Columns**.
This displays a list of the columns that can be hidden or shown.
2. Clear or select check boxes in this list, depending on which columns you want to hide or show.
3. Click anywhere outside of the list to dismiss the list.

To restore the default set of columns, click the grid  icon and then click **Reset Columns**.

17.4 See Also

- [Scheduling and Running Tasks](#)
- [Handling Task Errors](#)

18

Using the File Manager

The File Manager enables you to manage data files and schema files within the InterSystems® Data Fabric Studio™ file system. Specifically, each file-based [data source](#) defines and uses an associated set of directories within this file system. Via the File Manager, you can view the contents of those directories (other than the actual file contents), remove files, and upload new files.

Typically you use the File Manager to upload schemas and to upload data for test purposes. In production use, the recommendation is to create an automation that writes files to the applicable directories from the source system or systems. As a consequence, you might use the File Manager only during the early phases of your work with the system.

18.1 Standard Subdirectories for a File Source

Each file-based [data source](#) defines and uses its own directory within the file system, given the name that you provide when defining that data source. That directory includes the following standard subdirectories:

- **Samples**—The purpose of this subdirectory is to hold any sample files that define schemas. You can upload samples to this directory and then import them into the [Data Catalog](#).
These files remain in this directory until you delete them, if ever.
- **Source**—This is the input directory for the data source. When you run a [recipe](#) that loads data from this data source, the recipe looks for a file in this directory (and the file name must match the filename pattern specified by the associated schema; see [Editing a Schema](#)).

You can upload files to this directory or you can create an automation that writes files here from the appropriate source. You can manually delete files from this directory.

- **Work**—When a recipe is started, the input file is moved from **Source** to **Work**. This prevents the file from being accidentally processed by a staging activity in a different recipe.
- **Archive**—When a file has been successfully processed, it is moved from **Work** to **Archive** and is renamed with the timestamp appended to the original name.

(If a file is not successfully processed, it is moved back to the Source directory.)

18.2 Viewing the Directory Contents for a File Source

To view the directories associated with a file-based data source:

1. Click the File Manager  icon in the application menu.
2. For **Choose a File Directory Source**, select the data source of interest.

The system then displays a list of folders like this:



The folder names are always Archive, Samples, Source, and Work.


3. Click any folder name to display the current contents of that folder. For example:



4. Optionally edit **Max Number of Files to View**. This change immediately affects the display.

18.3 Uploading Files

To upload one or more files:

1. Click the File Manager  icon in the application menu.
2. For **Choose a File Directory Source**, select the data source of interest.
3. Select either the **Samples** or the **Source** directory, depending on where you want to load this file.

The page then displays any existing files in that directory.

4. Click **Upload Files**.

This displays a new area on the right, the *upload area*.


5. Perform the following steps within the upload area:

- a. Drag and drop one or more files from your local computer to the upload area.

Or click the upload area and then browse to the directory that contains the files. Then select one or more files.

As you do this, the upload area displays a list of files that can be uploaded.

Note that you cannot upload a file larger than 500 MB.



- b. Optionally, if you have selected files that you do not want to upload, click the Remove  icon next to those files. Or click **Clear Files**.
- c. Click **Upload File(s)** in the upload area.

Now the files are uploaded and are displayed in the list of contents for the given directory. The upload area is reset and shows no files.

Note that the new files overwrite any existing files that have the same name. The upload process does not rename files. Also when you upload a new copy of a schema file, that does not automatically reimport the schema. When you are ready to reimport the schema, do that as described in [Reimporting a Schema](#).

18.4 Deleting a File

To delete a file:

1. Click the File Manager  icon in the application menu.
2. For **Choose a File Directory Source**, select the data source of interest.
3. Select a directory name.
4. Click the Delete  icon next to the file.
5. Click **Remove** to confirm the action.

18.5 See Also

- [Importing Schemas](#)
- [Defining Recipes](#)


19

Exporting and Importing Configurations

For your solution, InterSystems (or other solution partner) sets up and maintains multiple parallel [environments](#) so that development and testing activities can safely occur without disturbing mission-critical activities. To simplify the process of keeping the environments synchronized, InterSystems® Data Fabric Studio™ provides options for exporting and importing configurations. Here the term *configuration* refers to any of the items that you configure within the solution, including schemas, recipes, users, and roles.

19.1 Exporting Configurations

To export one or more configurations:


1. Click the Configuration Management  icon in the application menu.
2. Click **Export Configurations**.
3. Click the check box next to each configuration that you want to export.
4. Click **Export Selected**.

The system creates a ZIP file that contains all the selected configurations and it downloads that file to the download location specified by the browser. The name of the file has the form `BundleZip<datetime>.zip`.

The ZIP file contains one file for each configuration, and for those files, the filename format is *itemname.itemtype*, for example: `MyTestRecipe.TotalViewRecipe`.

19.2 Importing Configurations

To import one or more configurations:

1. Create a ZIP file that contains the configuration files.
2. Click the Configuration Management  icon in the application menu.
3. Click **Import Configurations**.
4. Click in the upload area and navigate to your file.
5. Click **Import All**.

The system displays a dialog box indicating whether the import was successful, along with information about any errors.

6. Click **Close**.

19.3 See Also

- [About Your Solution](#)