

# Data Load using the Reltio Console UI

Data Sheet | LumenData

Both Entities and relationships can be loaded with simple, nested, and reference attributes using a data loader. The format of the file loading can be CSV, Excel, or JSON. For CSV files up to 33GB and Excel, 100k rows can be loaded. Source files can be from local workstations, AWS, GCP, Azure, or SFTP. Through the Console data loader UI, large amounts of data can be loaded into the tenant.

## It supports the following options:

- Data loading can be scheduled.
- Optimizes the data load process and automates queue management.
- Minimizes consumption of tenant quota.
- Provides metrics about a job.

## Data Loader loads data through the following steps:

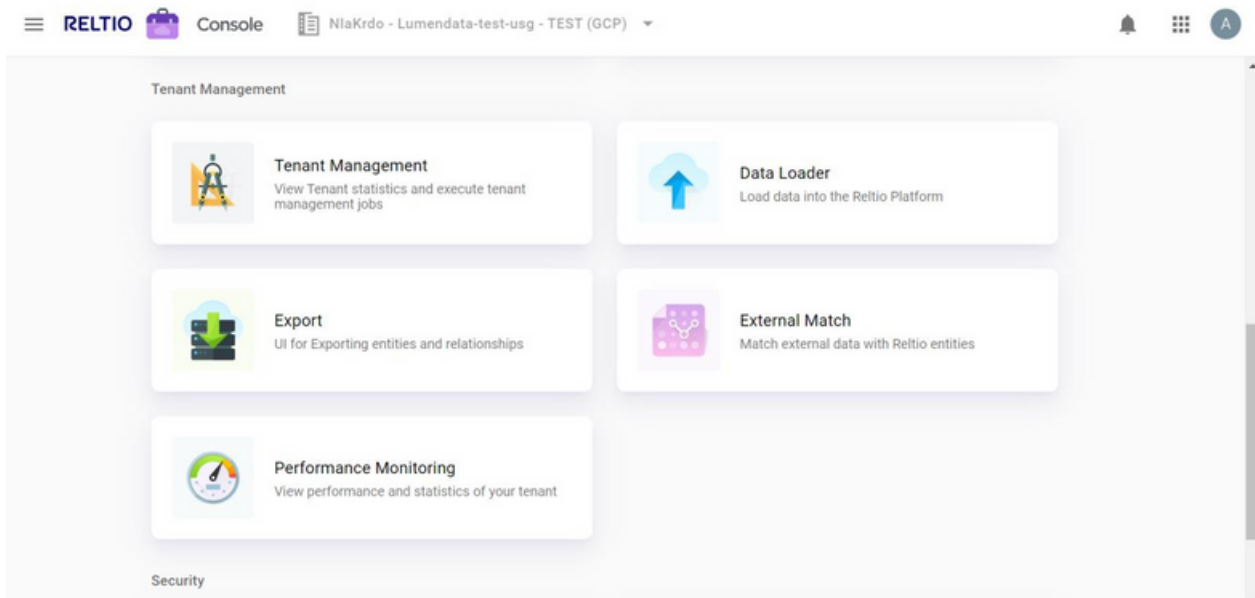
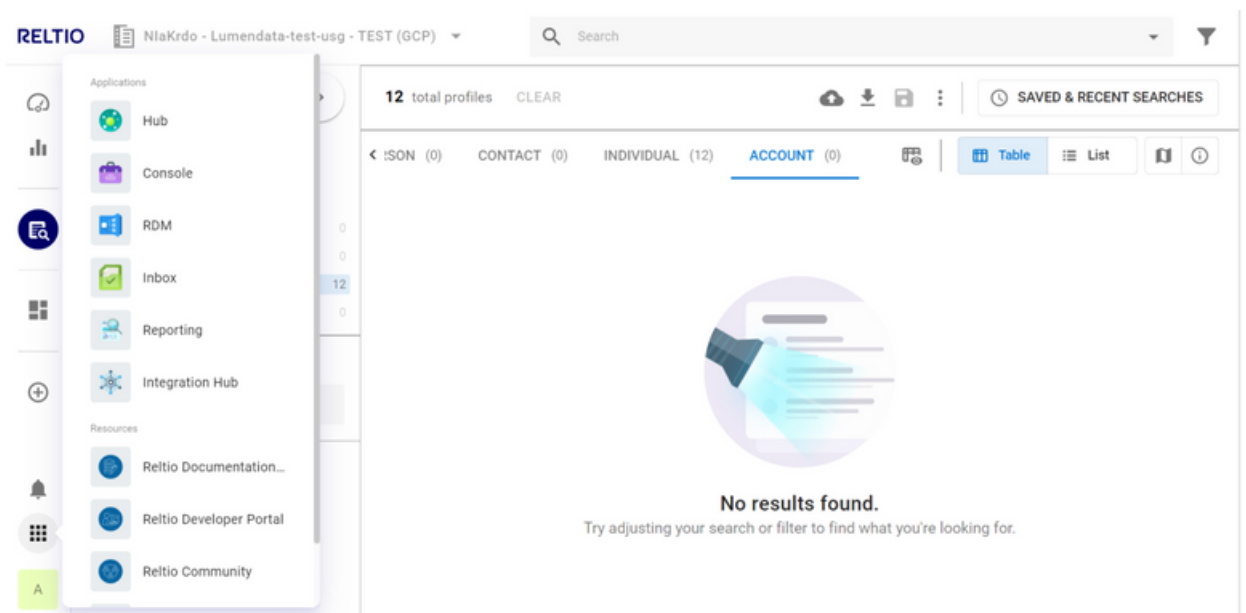
1. Create a new job or duplicate an existing job.
2. Select the input file from the designated source system.
3. Define new mappings or use an existing mapping to map file columns to the entity attributes.

4. Specify the records update preference.
5. Analyze key metrics from the input file (optional) before proceeding with the data load.
6. Start the data load.
7. Review data load completion metrics.

The following steps include how to perform data loading through UI.

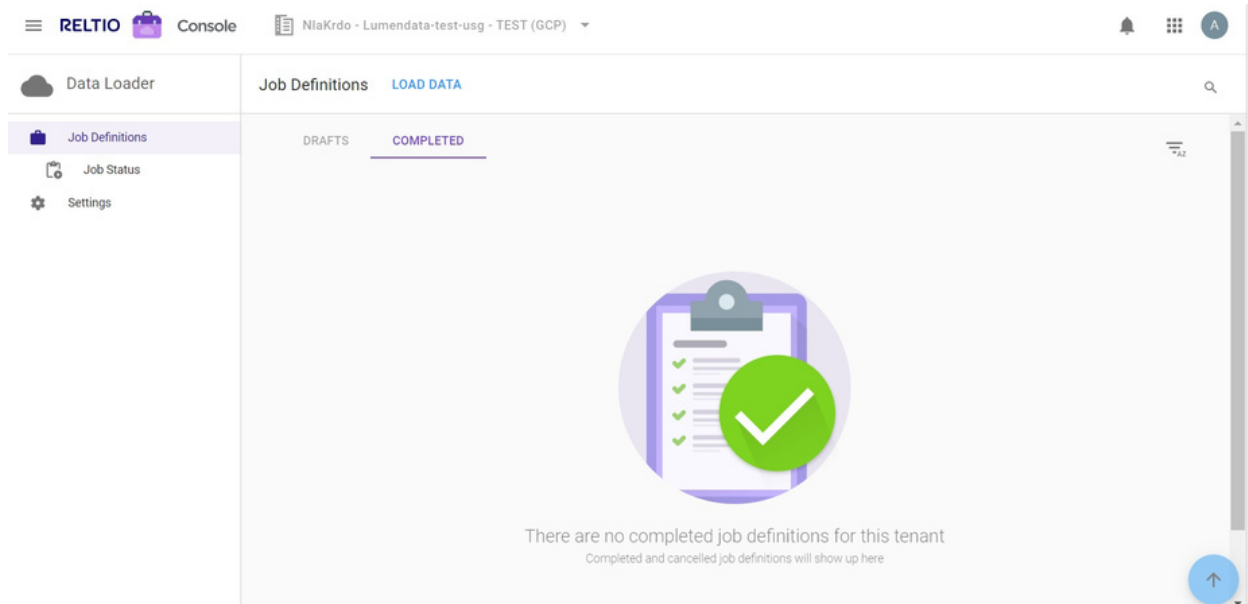
## 1. Create a Job

We can create a new job or duplicate an existing job. Go to the Multi Selector icon in the left corner of the tenant. Select the data loader from the Console.

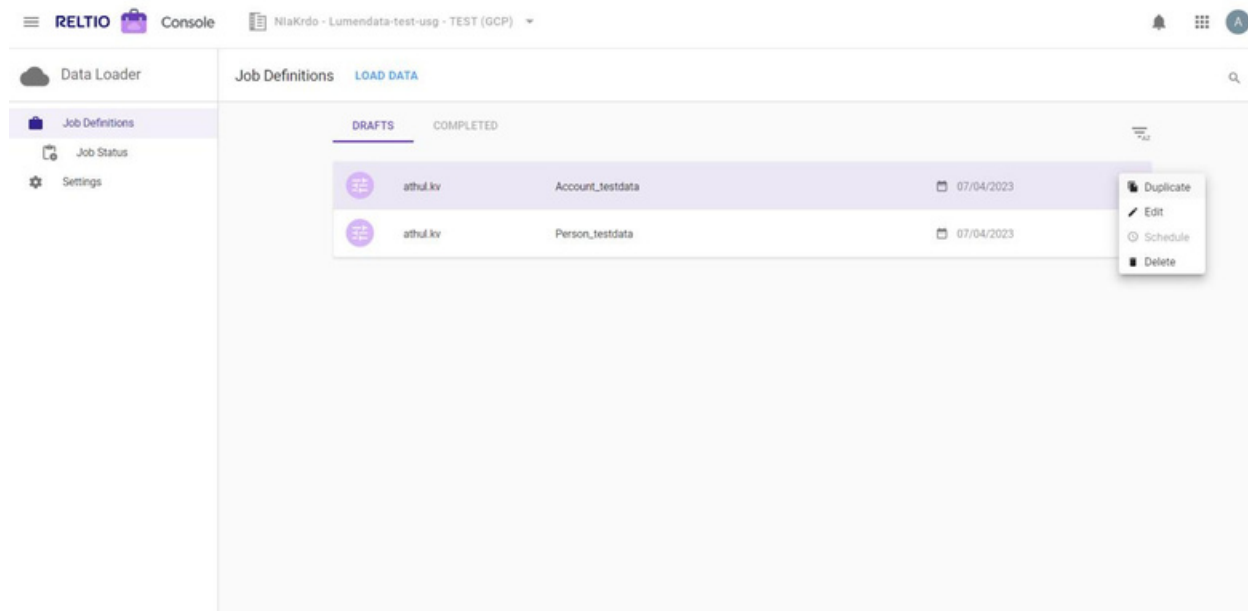


The first step in running the data loader is to create a job. This can be done by several methods.

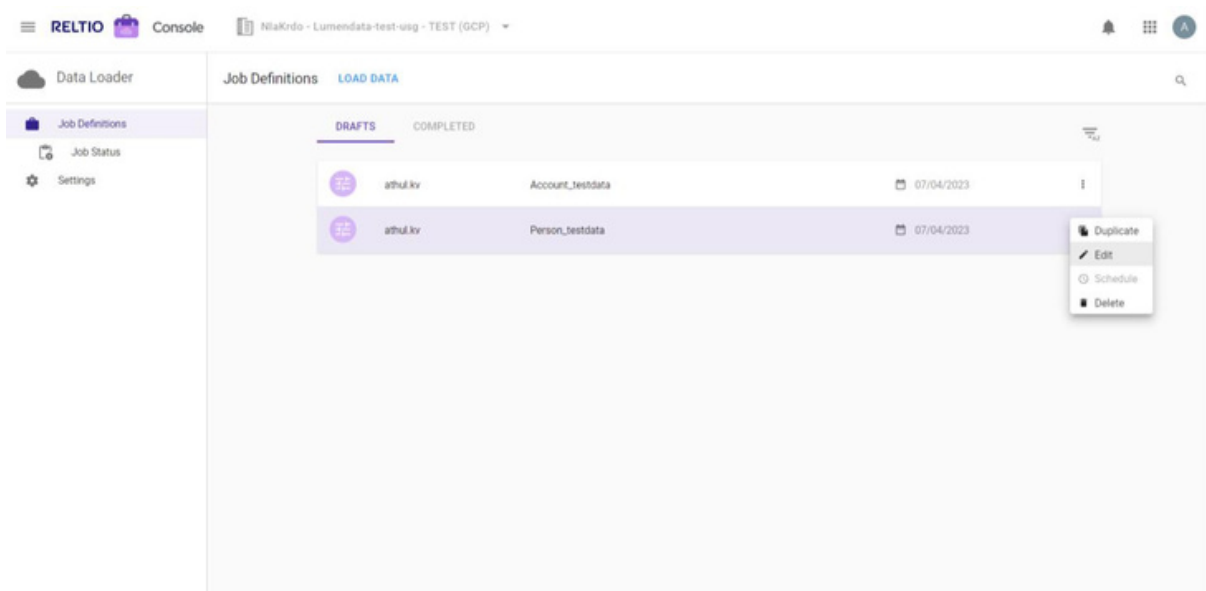
### 1.1 Create a new job by clicking **Load Data** followed by Entity and Relationship.



### 1.2 Create a new job using an existing job as a template using the **duplicate** option from the existing job.



### 1.3 Create a new job by re-executing the existing job and providing a new source file using the **edit** option in the job.



## 2. Specifying the source of the data

Data loader uses source data files from different locations.

- Local workstation
- AWS
- GCP
- Azure
- SFTP

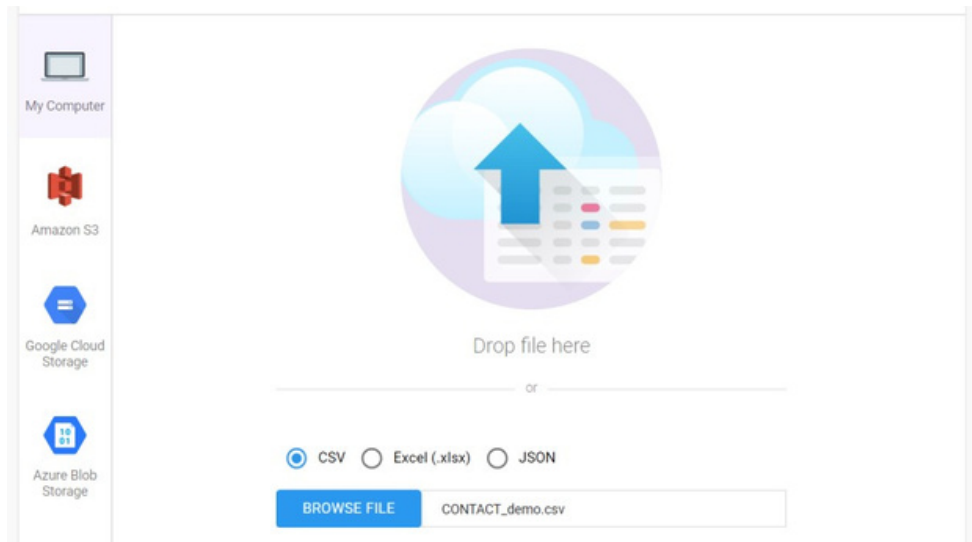
The AWS, GCP, Azure and SFTP source definitions have two groups of properties:

**i. Source-Specific properties:** These are different for each source and include the source credentials, host details, host directory, bucket, or container name that contains the source file(s).

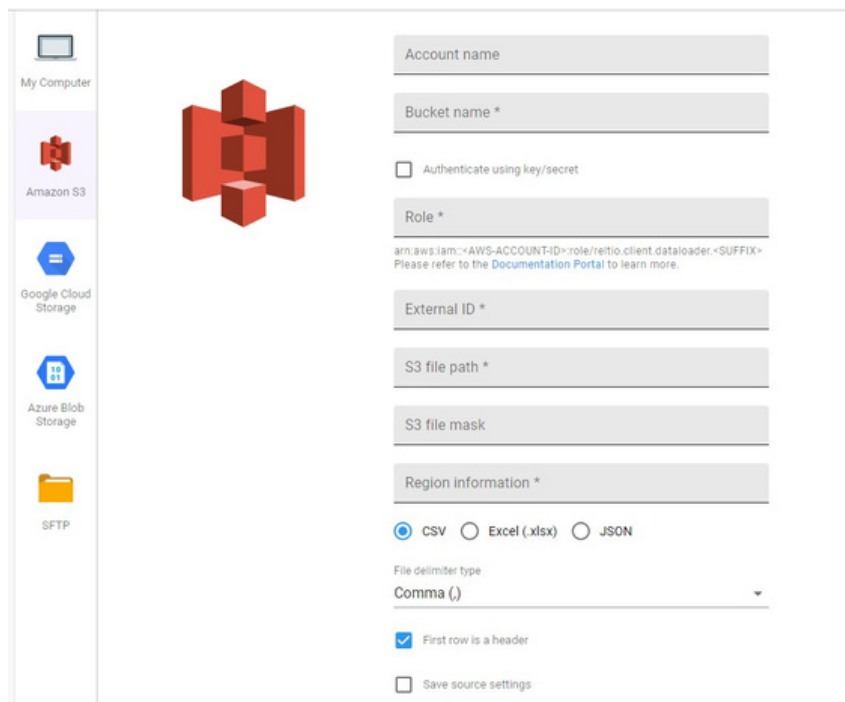
**ii. Data Loader Properties:** These are common to all the source definitions.

1. Account name: A name that you can specify to identify the job
2. File Mask: A prefix string to identify the files to be loaded from the source.
3. Save Source Settings: A flag to indicate whether the source details can be shared with other jobs or are private to this job.

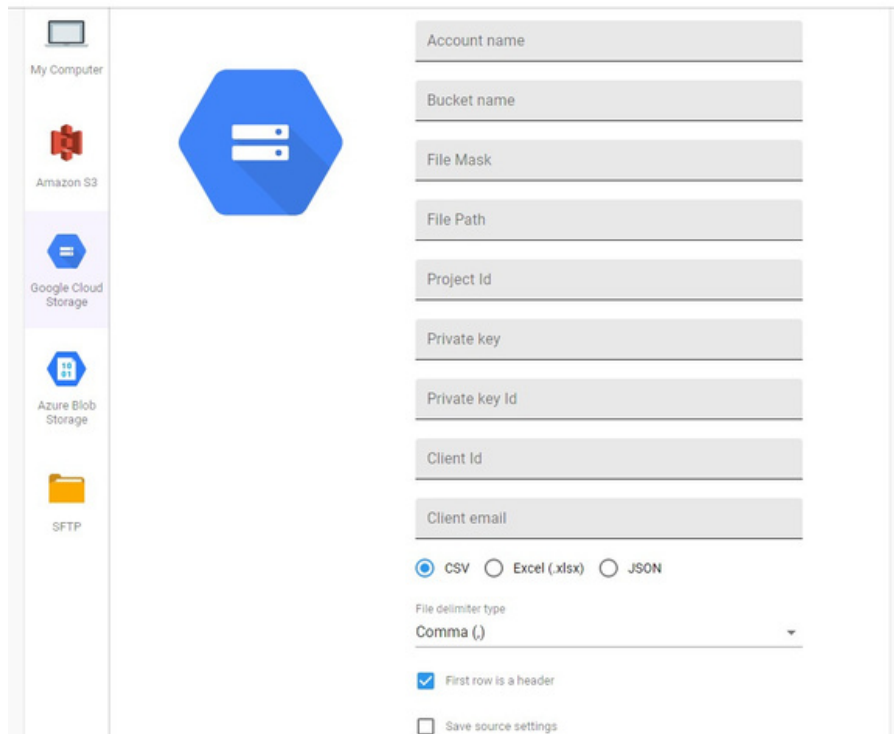
**2.1 Local workstation:** The Browse file option is used to select file from the local system.



**2.2 AWS:** While using AWS S3 as source five Parameters need to be specified. They are AWS Key, AWS Secret, AWS Region, S3 Bucket Name, and S3 path.

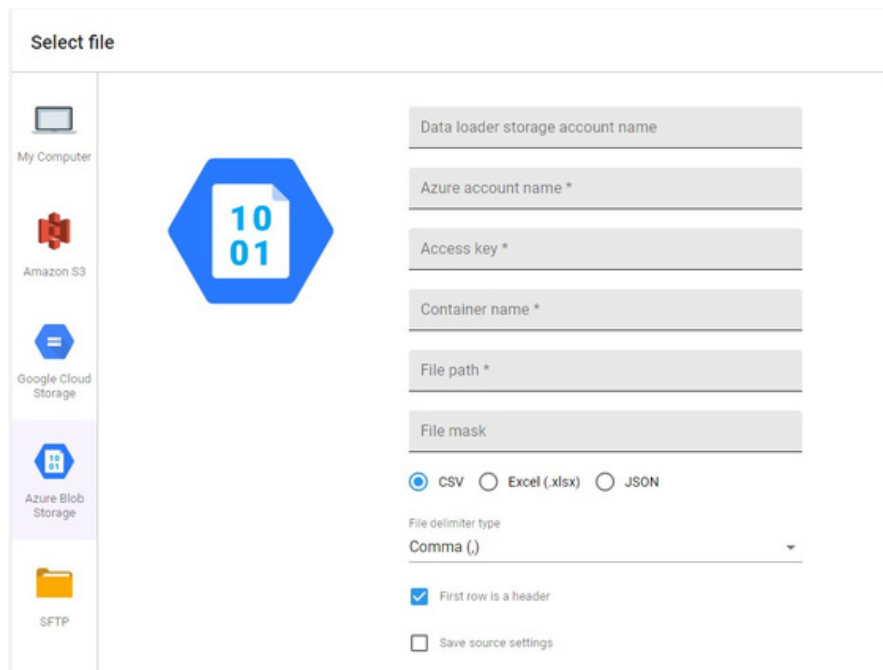


**2.3 GCP:** While using GCP as source, seven Parameters need to be specified. They are Project ID, Private key, Private Key ID, Client ID, Client Email, Bucket Name, and file path.



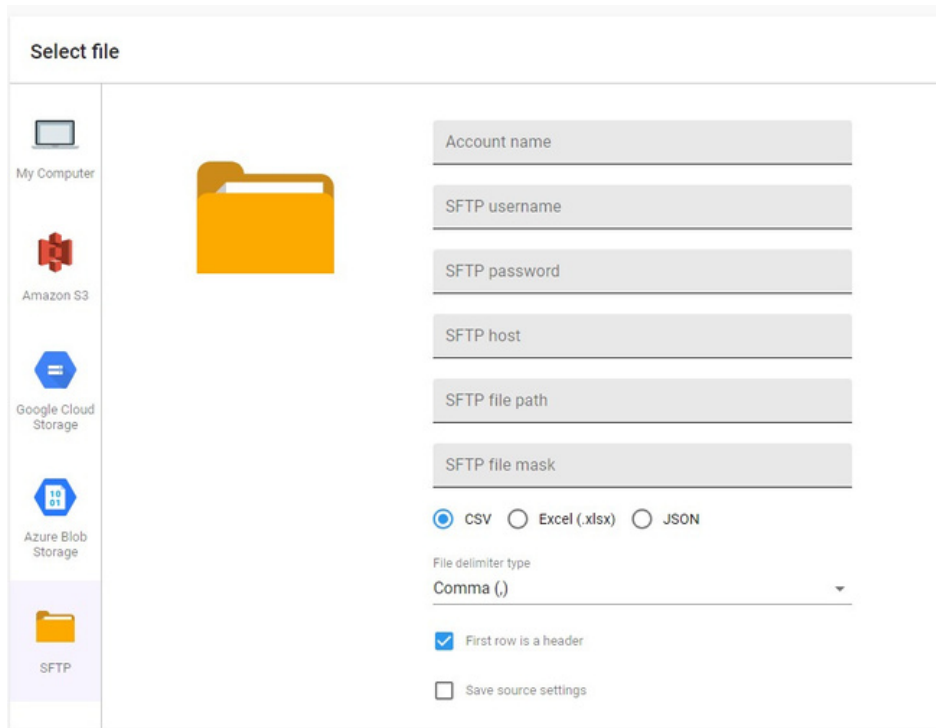
The screenshot shows a configuration window for Google Cloud Storage. On the left, there is a sidebar with icons for 'My Computer', 'Amazon S3', 'Google Cloud Storage' (highlighted), 'Azure Blob Storage', and 'SFTP'. The main area features a large blue hexagonal icon with a white document symbol. To the right, there are several input fields: 'Account name', 'Bucket name', 'File Mask', 'File Path', 'Project Id', 'Private key', 'Private key Id', 'Client Id', and 'Client email'. Below these fields are radio buttons for file formats: 'CSV' (selected), 'Excel (.xlsx)', and 'JSON'. A 'File delimiter type' dropdown menu is set to 'Comma (,)', and a checked checkbox indicates 'First row is a header'. An unchecked checkbox at the bottom is labeled 'Save source settings'.

**2.4 Azure:** While using Azure as source, four Parameters need to be specified. They are Azure Account name, Access Key, Container Name, and file path.



The screenshot shows a configuration window for Azure Blob Storage. The sidebar on the left includes 'My Computer', 'Amazon S3', 'Google Cloud Storage', 'Azure Blob Storage' (highlighted), and 'SFTP'. The main area contains a large blue hexagonal icon with a white document symbol and the numbers '10 01'. The right side of the window has input fields for 'Data loader storage account name', 'Azure account name \*', 'Access key \*', 'Container name \*', and 'File path \*'. Below these are fields for 'File mask', radio buttons for file formats ('CSV' selected, 'Excel (.xlsx)', 'JSON'), a 'File delimiter type' dropdown set to 'Comma (,)', a checked checkbox for 'First row is a header', and an unchecked checkbox for 'Save source settings'.

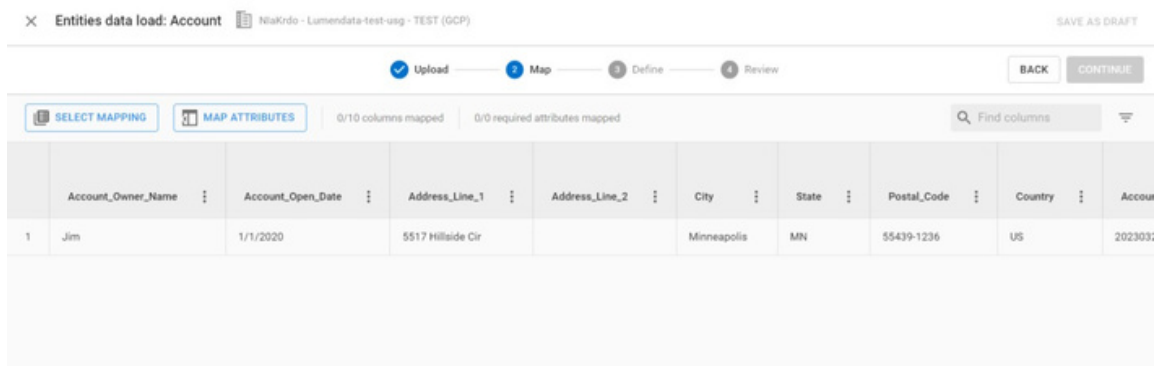
**2.5 SFTP:** While using SFTP as source, four Parameters need to be specified. They are Username, Password, Host URL, and Source file sub-directory.



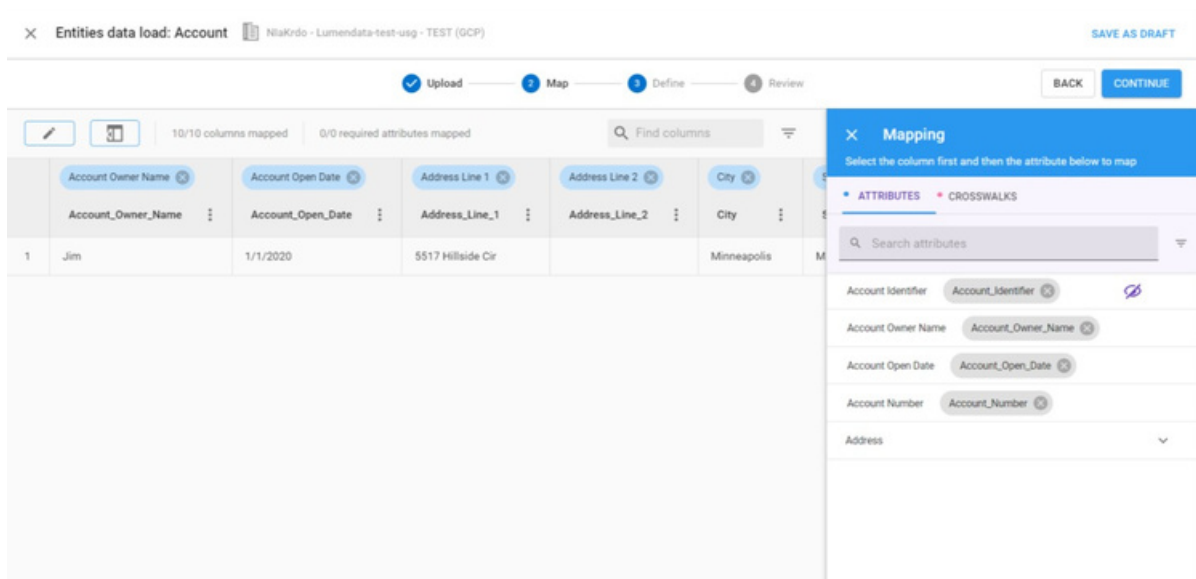
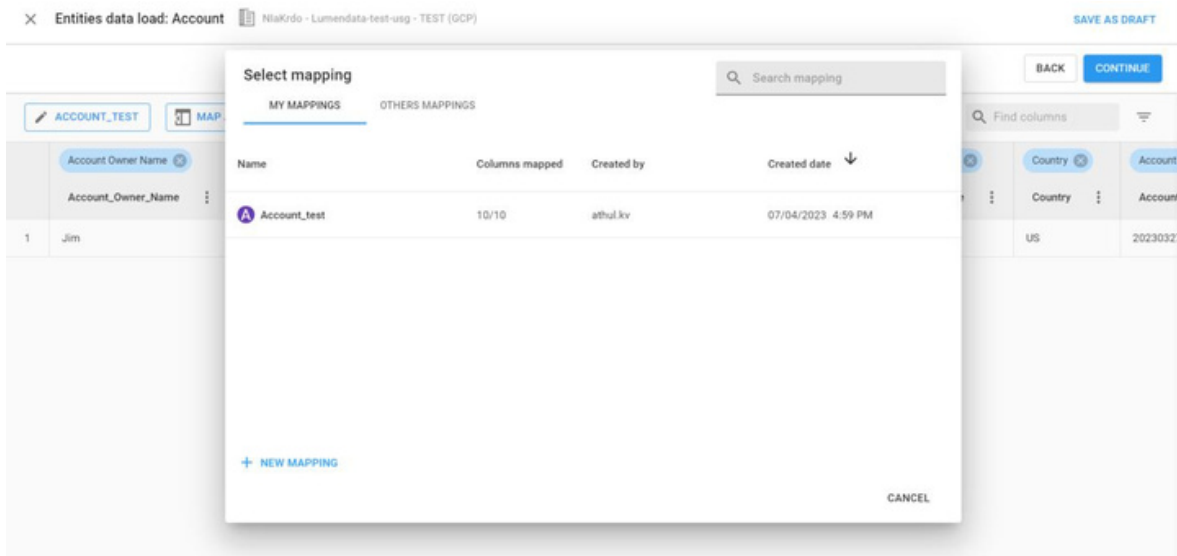
### 3. Mapping

Creating a mapping tells Reltio how to utilize a column in a source file to map the appropriate attribute. The Reltio Console Data Loader allows mappings to simple, nested, and reference entity attributes in a Reltio tenant. In the case of Relationships, simple and nested attributes can be mapped.

Users can create a new mapping or existing mapping.



	Account_Owner_Name	Account_Open_Date	Address_Line_1	Address_Line_2	City	State	Postal_Code	Country	Account
1	Jim	1/1/2020	5517 Hillside Cir		Minneapolis	MN	55439-1236	US	2023031



## 4. Schedule and run the job

If data is loaded from the local workstation, then we can't schedule job. If source files are stored in AWS, GCP, and SFTP we can schedule jobs. For that, the user need to specify job actions. They are:

- Job Definition name
- Select update: It can be a full or partial update.
- Enable/disable Execute LCA



Entities data load: Account | NIKIRIO - Lumendata-test-usg - TEST (GCP) | SAVE AS DRAFT

Upload | Map | **Define** | Review

BACK CONTINUE

**Select update**

Full Update  
Add or overwrites all attributes - both explicitly mapped and unmapped. You will lose existing values for unmapped attributes when they are overwritten with empty values.

Partial Update  
Add or overwrites only the explicitly mapped attributes. Leaves existing attribute values unchanged.

**Advance settings**

**Partial override for reference**  
If selected, platform overwrites all attributes, but for references, we do not touch any attributes not involved with referencedAttributes.

**Update crosswalks singleAttributeUpdateDates property**  
If selected, in case of an override of an Entity, when a match with an existing entity is found, the singleAttributeUpdateDates property of crosswalks is updated for changed and new attributes.

**Execute LCA**  
If selected, Lifecycle Actions(LCA) will be triggered during this request. Default is true.

**Enable Data Change Request (DCR)**  
Records loaded will go through DCR Workflow process and will not be loaded directly to the tenant.

**Preserve hidden**  
If selected, API will not create a DCR to delete hidden attributes if they are not specified in the request.

**Recurring Job Definition**  
Turn on if you want this job to repeat.

**Recurring Job Definition:** Users can run a job at a specified interval of time, while the data is loaded from AWS, GCP or SFTP.

**Recurring Job Definition**

Turn on if you want this job to repeat

The system will search for a new file in the folder <insert folder name> according to the schedule.  
Data load operations will only run if a new file was identified.

**Start Date**  Now

**Repeat Every** 1

**Repeat On**  Mon  Tue  Wed  Thu  Fri  Sat  Sun

**End On**  Never  On

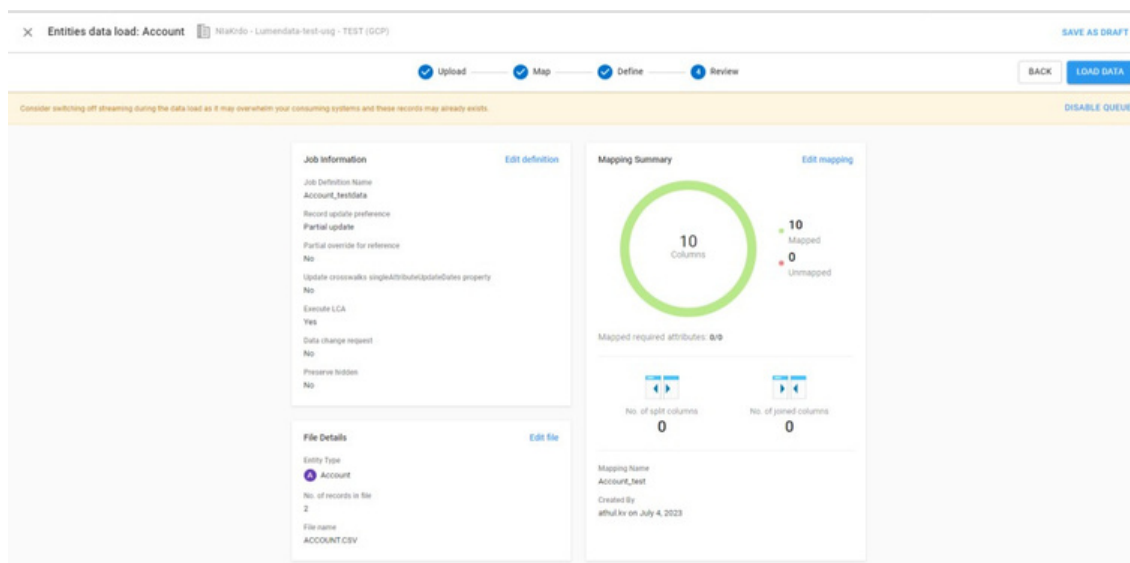
Summary : Repeats every hour

## 5. Review Job details

After completing job definition, we will get a summary of jobs. It includes:

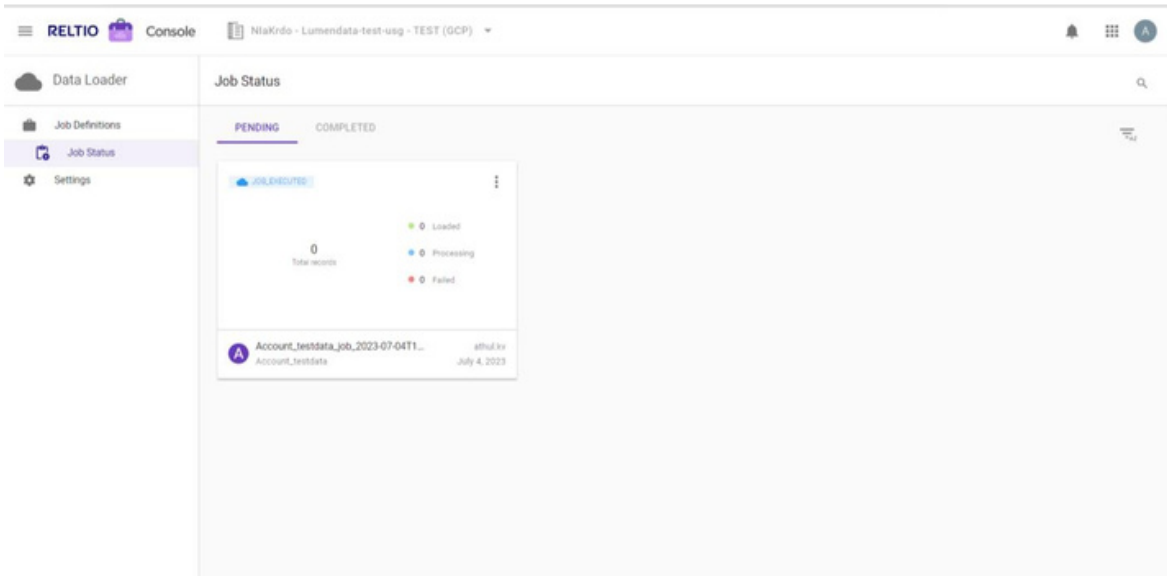
- **Disable Queue:** Matrices about the job. We can disable the series of events if we want to.

- File Details: We can return to the file upload screen and able to edit.
  - Entity Type/ Relationship Type
  - Number of Records in File
  - File selected from which location.
- Mapping summary: Users can go back to the mapping screen and edit mapping using this.
  - Mapping:
  - Map settings
  - Mapped columns
- Job definition: User can edit job definitions from the job definition screen.

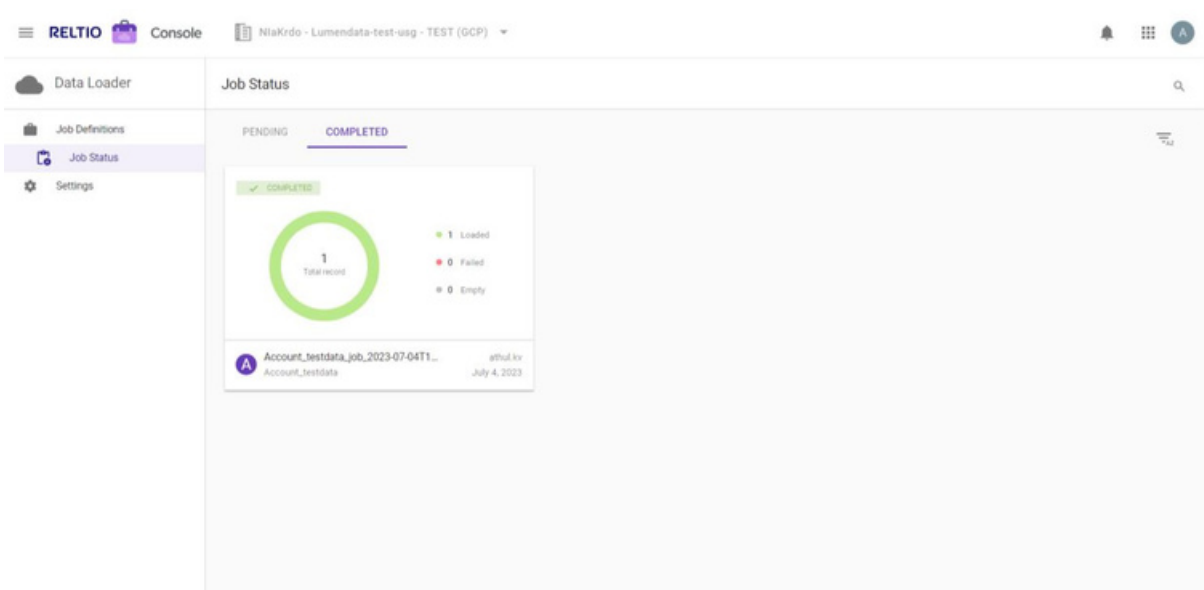


## 6. Monitor Job

Job status is used to monitor jobs once we have submitted the job. The pending tab displays jobs that are still being processed. It includes the current state of the job along with the number of records loaded, processed, and failed.



Once the job is finished, jobs are moved to the Completed tab. It includes job status as completed along with the number of records loaded and failed. In the completed tab jobs are displayed in green or red color. Green indicates the job completed successfully. If any of the records didn't process and the status of the job is error and red. We can download details of errors.



## ABOUT LUMENDATA:

LumenData is a leading provider of Enterprise Data Management, Cloud & Analytics solutions. We help businesses navigate their data visualization and analytics anxieties and enable them to accelerate their innovation journeys. Founded in 2008, with locations in multiple countries, LumenData is privileged to serve over 100 leading companies, including KwikTrip, Versant Health, US Food & Drug Administration, US Department of Labor, Cummins Engine, BCG, and others. LumenData is SOC2 certified and has instituted extensive controls to protect client data, including adherence to GDPR and CCPA regulations.

Get in touch to discuss how we can facilitate data-driven transformation for your organization.

---

## MEET OUR AUTHORS

---



**Nithin**

Associate Consultant



**Athira KM**

Associate Consultant



**Mohd Imran**

Senior Consultant