DATA WAREHOUSING WITH IBM CLOUD DB2 WAREHOUSE

# Abstract

Building the data warehouse using IBM cloud Db2 warehouse. Defining the schema and structure of the data warehouse tables. Identify data sources and design a strategy to integrate them into the data warehouse.

## **Data Modelling**

Before building a building, we first need to create its design and make a model. In the same way, to create a data warehouse, we need to design it first using data warehouse modelling tools and techniques. We do this to represent the data in the real world and see how business concepts relate. Data warehouse modelling is the process of designing the summarized information into a schema.

## **Schema**

Schema means the logical description of the entire database. It gives us a brief idea about the link between different database tables through keys and values. A data warehouse also has a schema like that of a database. In database modelling, we use the relational model schema. Whereas in the data warehouse, we use modelling Star, Snowflake, and Galaxy schema.

**Steps to build a database with CSV files**

1. Upload and convert multiple CSV files as a resource
2. Import your resource to a data workspace project
3. Transform your tables into an appropriate format and schema
4. Add new CSV files to update your table

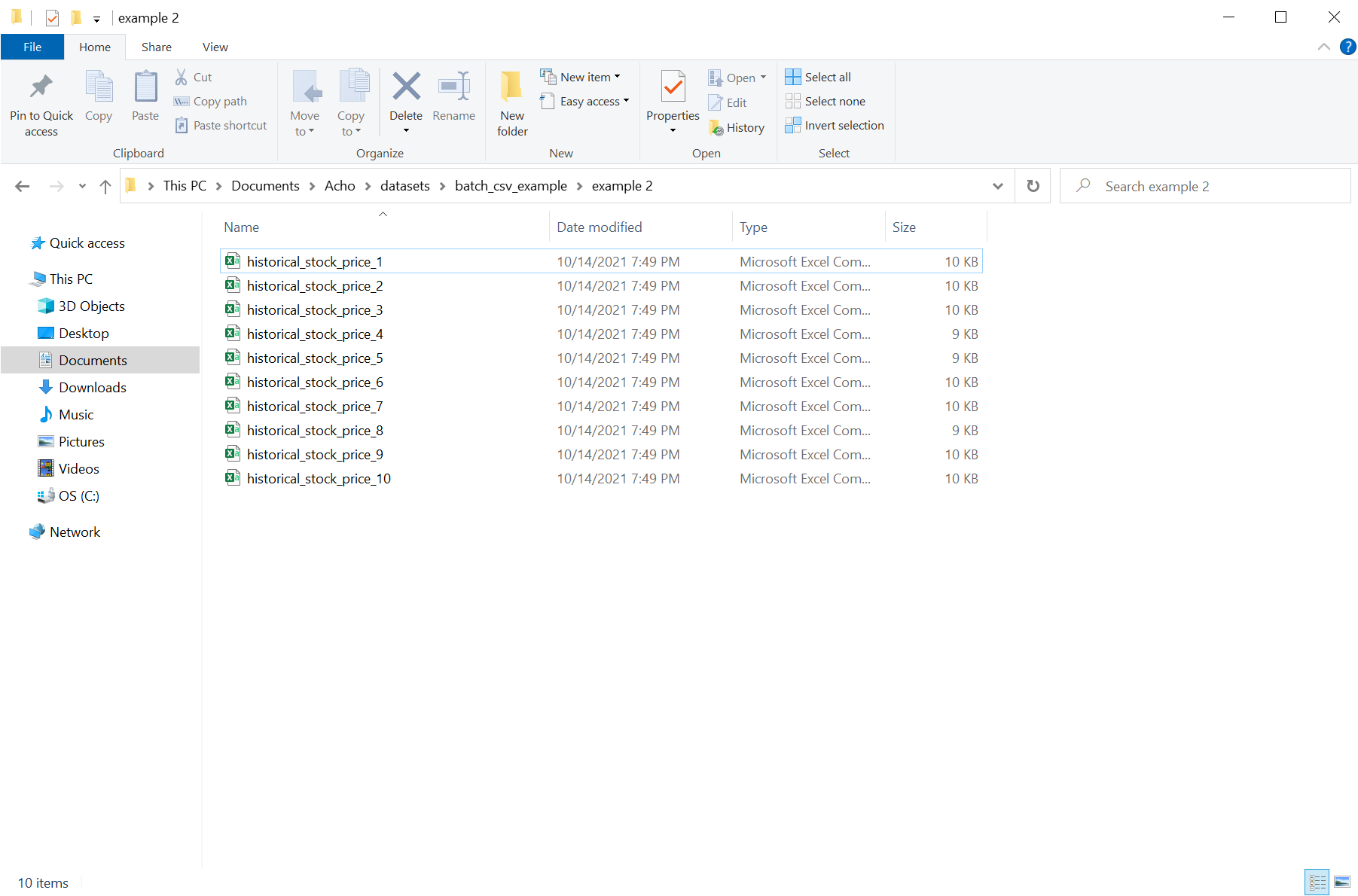
## **Build a database with CSV files**

Building a robust relational database takes time and effort. But you can do it within minutes using Acho. I will now show you how you can convert your CSV files into a SQL database in three simple steps.

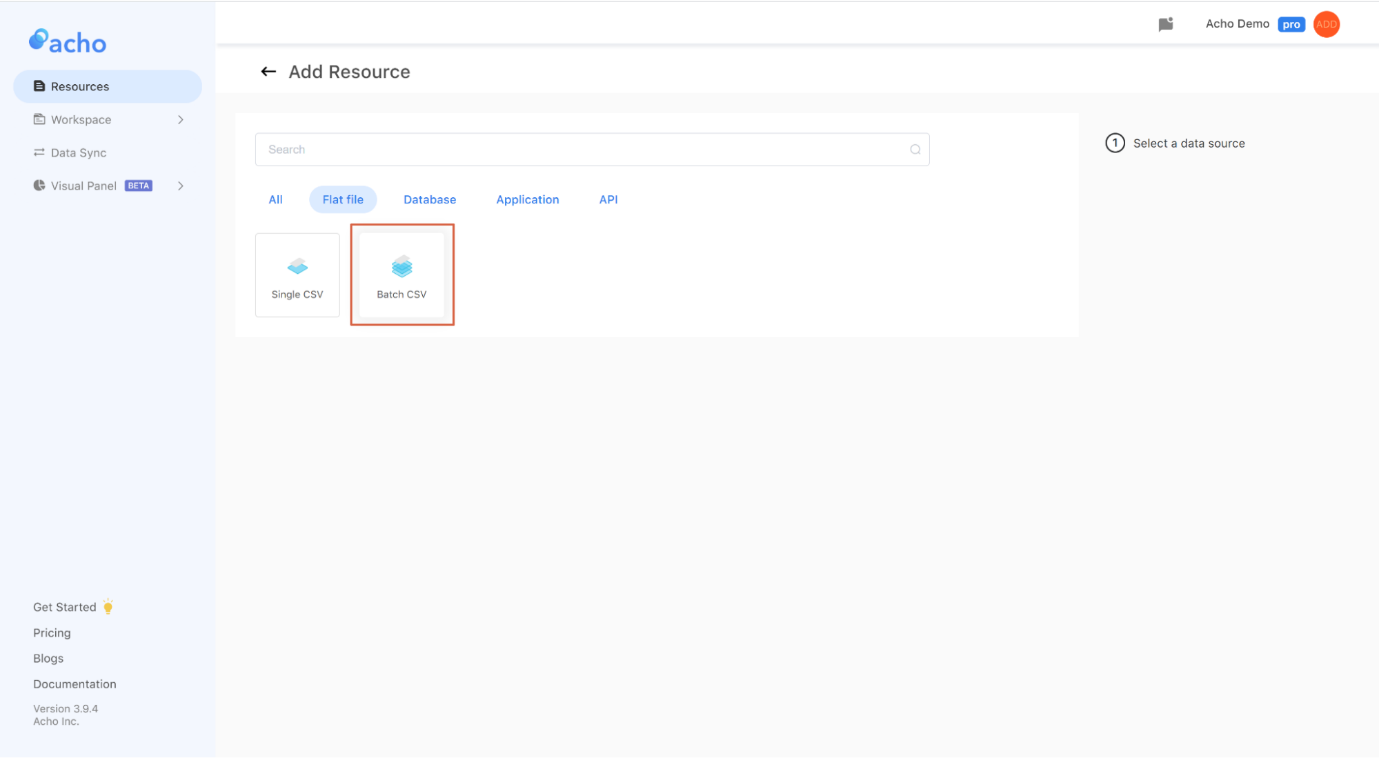
**1. Upload and convert multiple CSV files as a resource**

* Upload multiple CSV files as a Batch CSV resource on Acho
* Make sure the column names in all files are the same

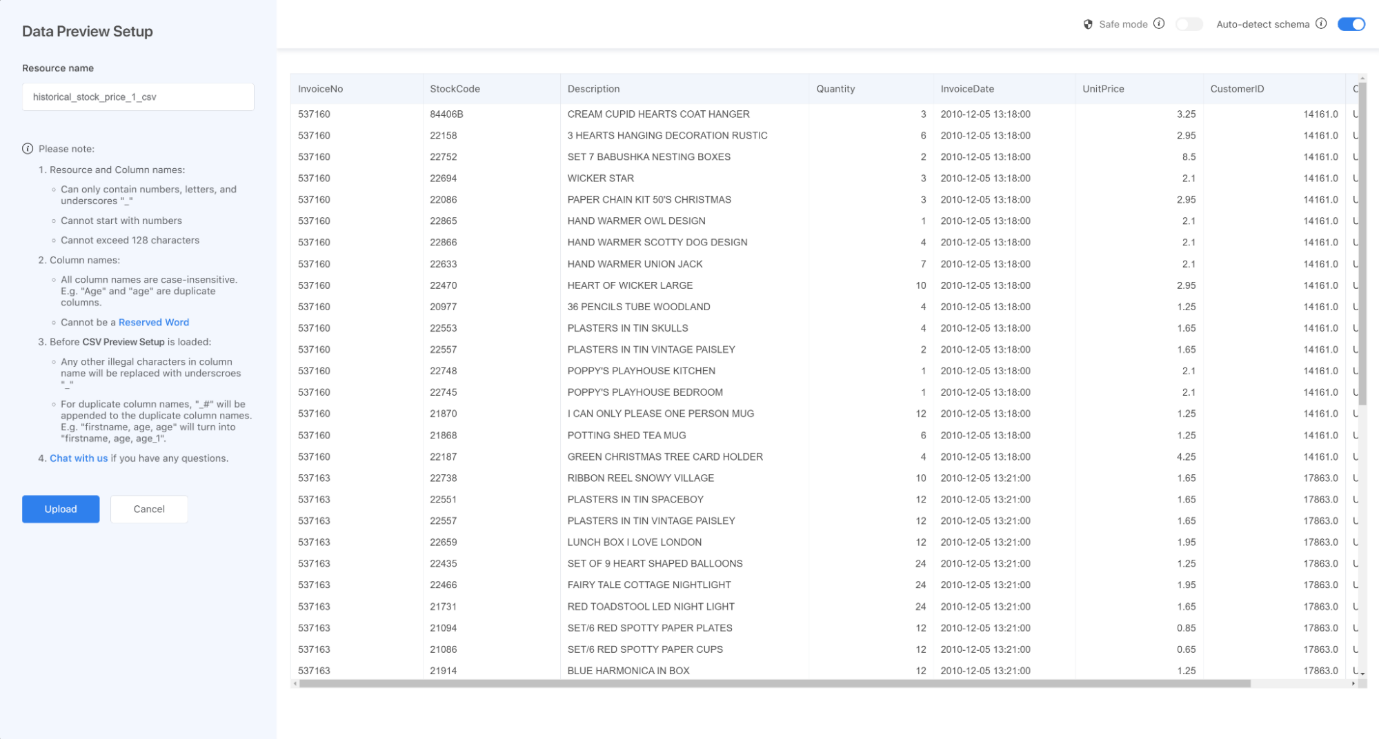
First, import your files to Acho Studio as a resource. Here I have a folder called “acho\_export\_historical\_stock\_price”. It contains some historical stock price records for 5 years. There are 10 files in this folder. All of them have the same column name.

 Build a Database with CSV files on Acho

Then go on to Acho Studio’s resource page and select “Batch CSV” for importing and converting to your resource. Be aware that only .csv, .tsv, and .txt files are supported at the moment.

 Upload CSV files as a Batch CSV resource

If they all have the same column name and correct values for each column, you should see a preview page with a sample of your files like this. (Note: If all of your files don’t have a header row, remember to switch to a “Safe Mode” to add a header)

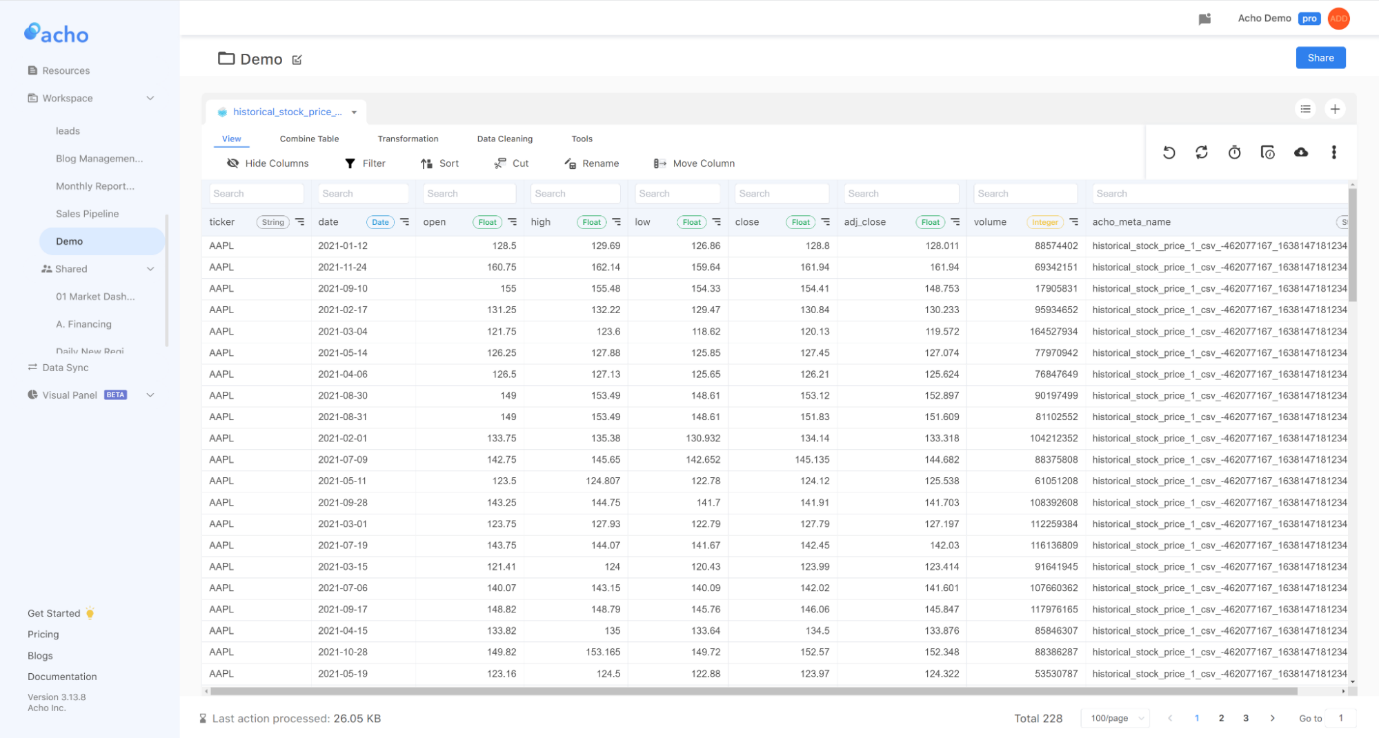
 Batch CSV file preview

**2. Import your resource to a project**

* Create a project based on the resource
* Check the total row number to make sure there’s no missing data

Once you have successfully uploaded the .csv batch files as a resource, you can now create a project to access them.

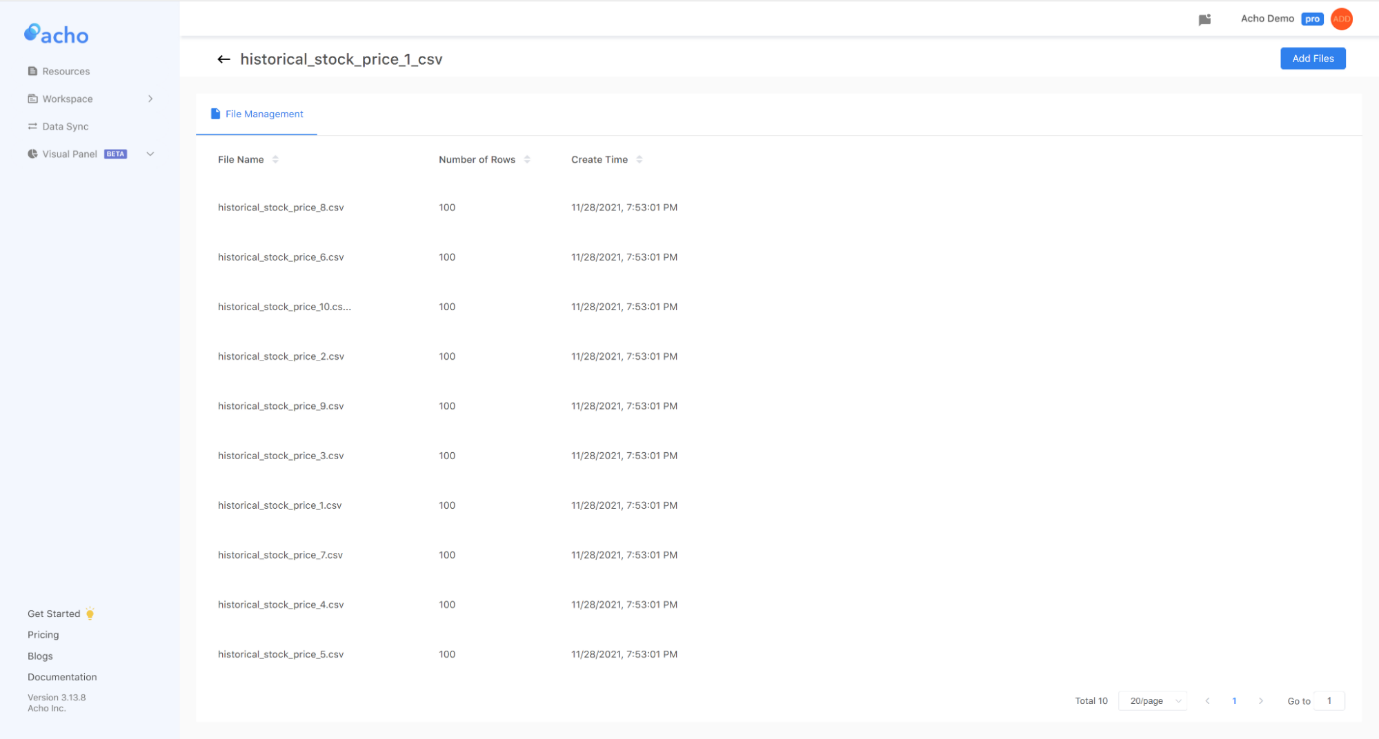
As you can see, all records from the 10 .csv files are now stored in one table.

 Acho Project for CSV Files

**3. Update your table with new CSV files**

* Click  “File Management” button next to your Batch CSV resource
* Go to your project
* Click “Sync with the Batch CSV resource”

It is simple to upload your table with new CSV files. You can simply go to the resource page and click the “File Management'' button next to your Batch CSV resource. Here you can upload the new files and the system will automatically merge the new files into the current resource and aggregate all the records based on the column names. There is no limit on how many rows you can add to the resource. All the records are stored in a relational database on the cloud, so it’s pretty easy to transform, cleanse and update them. (Notice that all news files should have the same column names with the current resource.)

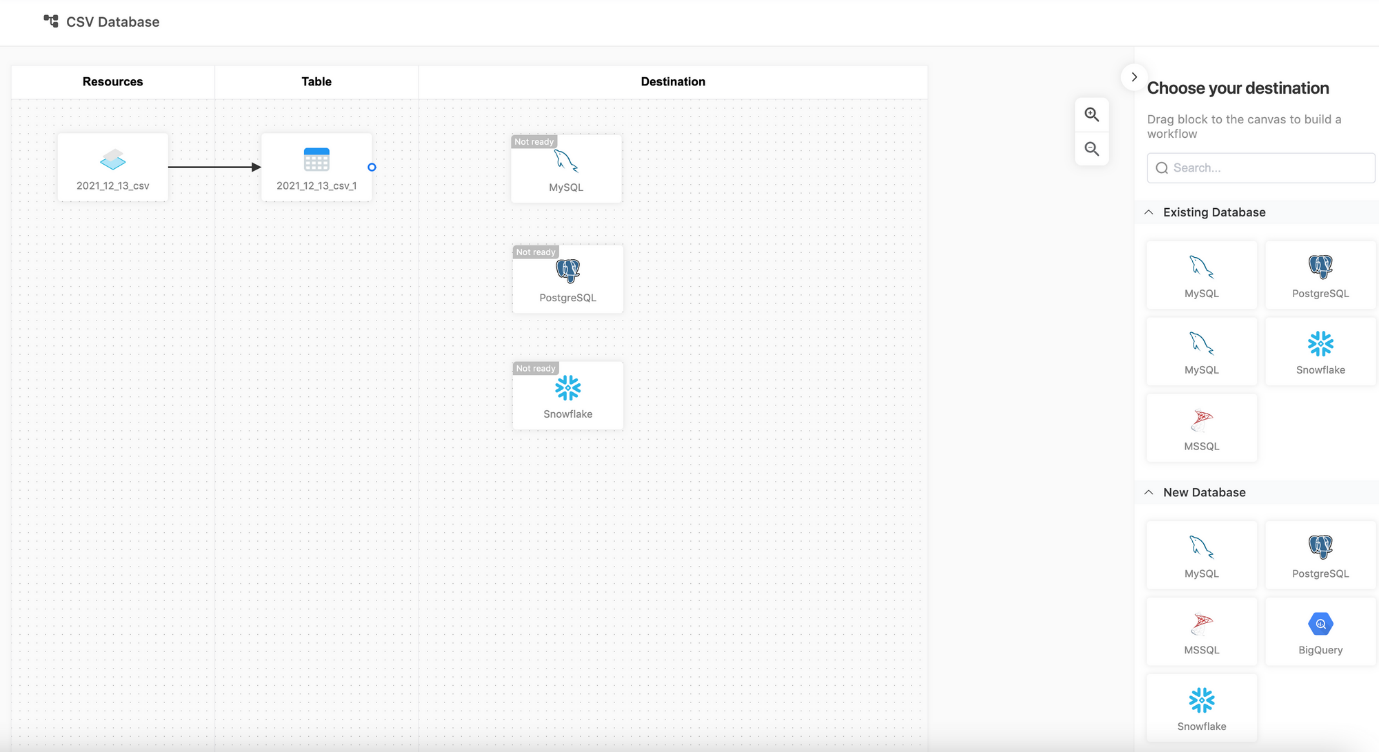
 Update Batch CSV

Then, go to the project where you imported the Batch CSV resource, and click the “Sync with the Batch CSV resource” button. Now, the system will retrieve the latest records from the resource.

If your new data files have different column names or you want to expand your tables horizontally, you can upload the new file as a new resource and check out this [article](https://acho.io/blogs/how-to-combine-dozens-of-big-datasets-together-in-a-minute-2?ref=ghost.acho.io) to merge them into your current table by using the “Union” or “Join” action.

**4. Send the data to a production database**

Once the data is imported and built out in the project table, you can use the "Workflow" feature to export the table to a production database such as MySQL, PostgreSQL, Big Query, and more. By verifying the credentials, you can write the table data to the database automatically.



CSV to Database on Acho

How to integrate csv files into data warehouse

Here are a few step-by-step guides that show you how:

* [How To Setup A Batch Data Pipeline For CSV Files To Redshift, Redshift Spectrum, Athena or BigQuery](https://blog.openbridge.com/how-to-setup-a-batch-data-pipeline-for-csv-files-8c4d0cd7394b)

Automated workflow provides you with the ability to bulk load data from external services like[Salesforce Marketing Cloud](https://blog.openbridge.com/export-tracking-data-from-salesforce-marketing-cloud-8a0a4c1f37dc), ERP, Adobe, and many other sources.

It is simple enough for non-technical users. You can also manually load data in an ad hoc manner with tools like FileZilla, Transmit, Cyber Duck…

* [FileZilla: 3 Simple Steps To Load Data to Big Query, Amazon Redshift, Redshift Spectrum or AWS Athena](https://blog.openbridge.com/filezilla-3-simple-steps-to-load-data-to-bigquery-amazon-redshift-redshift-spectrum-or-aws-athena-95b29a71a33c)

## **The advantages of building**

First, relational databases enable you to see the relationship between different parts of your data more easily.

Second, relational databases offer superior data management with greater consistency and flexibility.

Third, relational databases solve the problem of data redundancy.