

**Project:**

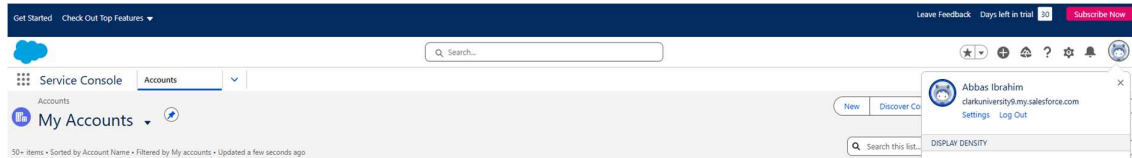
**How to Turn Data Into Insights From  
Ingestion**

## Deliverables:

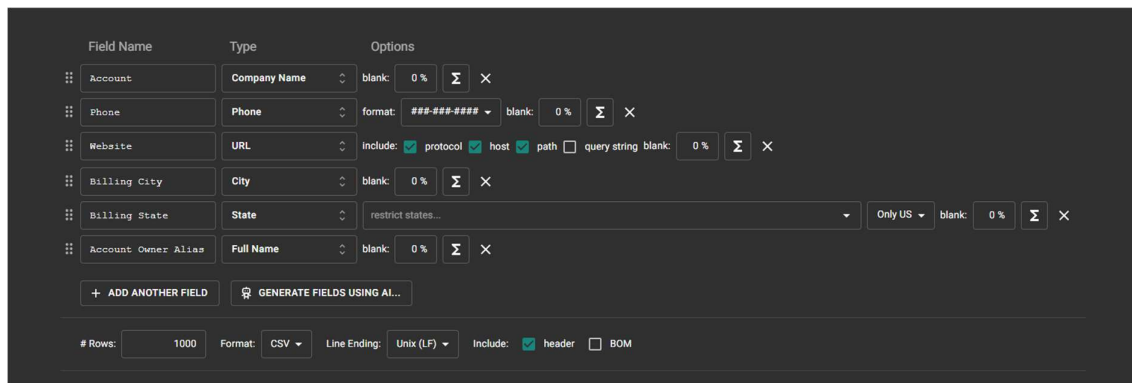
### Part 1: Generating Data and Configuring Your Tools

#### Data Generation :

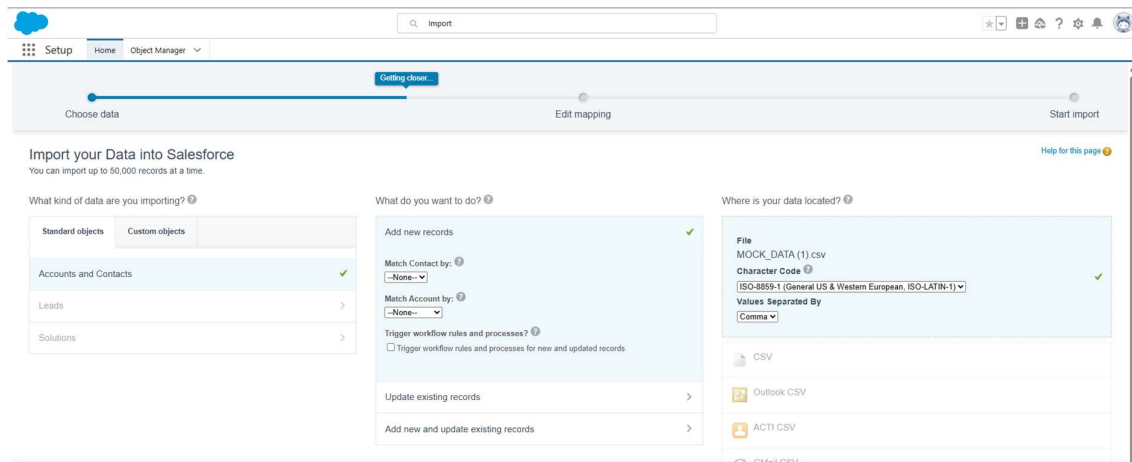
Created a salesforce enterprise account using the provided link



Generated a data set in Mockaroo using the below format with 1000 rows



Successfully imported generated data in Salesforce.



Mapped Salesforce Object to respective CSV header in order to import successfully without any errors.

SetupHomeObject Manager

Import

Almost done

Choose data

Edit mapping

Start import

Edit Field Mapping: Accounts and Contacts

Your file has been auto-mapped to existing Salesforce fields, but you can edit the mappings if you wish. Unmapped fields will not be imported.

Help for this page

Edit	Mapped Salesforce Object	CSV Header	Example	Example	Example
Change	Account: Account Name	Account	Yakjo	Jabbersphere	Yambee
Change	Account: Phone, Contact: Phone	Phone	646-712-8839	813-941-7756	616-394-8250
Change	Account: Website	Website	https://mac.com/heck	https://independent.co	http://cbc.ca/arsuscript/multi/et/1ac.png
Change	Account: Billing City	Billing City	New York City	Tampa	Grand Rapids
Change	Account: Billing State/Province	Billing State	New York	Florida	Michigan
Change	Contact: Name	Account Owner Alias	Gayler Fauguel	Alwyn Antosch	Clary Gregersen

CancelPreviousNext

Batches											
View Request	View Result	Batch ID	Start Time	End Time	Total Processing Time (ms)	API Active Processing Time (ms)	Apex Processing Time (ms)	Records Processed	Records Failed	Retry Count	Status
<a href="#">View Request</a>	<a href="#">View Result</a>	751Hu00001DKyGz	11/6/2024, 10:03 AM	11/6/2024, 10:03 AM	5768	5319	829	1,000	3	0	Completed

**noreply@salesforce.com** <noreply@salesforce.com>  
to me

Your Accounts and Contacts imports are complete. Here are your results:

Accounts Created: 997  
Accounts Updated: 0  
Accounts Ignored: 0 (We ignored updates that we couldn't match to an existing record.)  
Accounts Failed: 3 (We couldn't import these due to errors.)  
Processed job information for imported Accounts: <https://clarkuniversity9.my.salesforce.com/750Hu00000m2P6CIAU?fromEmail=1>

Contacts Created: 1,000  
Contacts Updated: 0  
Contacts Ignored: 0 (We ignored updates that we couldn't match to an existing record.)  
Contacts Failed: 0 (We couldn't import these due to errors.)  
Accounts/Contacts Rejected: 0 (We rejected duplicate rows.)  
Processed job information for imported Contacts: <https://clarkuniversity9.my.salesforce.com/750Hu00000m2P5yIAE?fromEmail=1>

\*\*\*

As soon as I tried importing the dataset in Salesforce based on mapping the objects, it got processed, and received a response mail from Salesforce confirming that all the contacts data got imported successfully but as for accounts 3 records didn't get imported successfully. While scrutinizing those three records, they were found to be duplicates and that's the reason they didn't get imported into Salesforce.

A	B	C	D	E	F	G
BillingCity	BillingState	Name	Phone	Website	Salesforce.com ID	Errors
Montgomery	Alabama	Yabox	(334) 555-2578	https://narod.ru/elit/ac.jpg		DUPLICATES_DETECTED:You are creating a duplicate record. We recommend you use an existing record instead.:-
Philadelphia	Pennsylvania	Babblestorm	(215) 378-7988	http://salon.com/ullamcorper.png		DUPLICATES_DETECTED:You are creating a duplicate record. We recommend you use an existing record instead.:-
Houston	Texas	Reallinks	(713) 807-7999	http://google.com.br/vulputate/justo/in/blandit/ultrices/enim/lorem.json		DUPLICATES_DETECTED:You are creating a duplicate record. We recommend you use an existing record instead.:-

After importing the generated dataset into my accounts, we could find our imported data below in the accounts tab

Accounts

My Accounts

50+ Items • Sorted by Website • Filtered by My accounts • Updated a few seconds ago

Search this list...

New

Import

Printable View

Account Name

Phone

Website

Billing City

Billing State/Province

Account Owner Alias

1

☐

Janyx

📞(727) 824-6139

https://zimbio.com/posuere/cubilia/curae.png

Saint Petersburg

Florida

Albra

2

☐

LiveZ

📞(904) 719-6420

https://zimbio.com/posuere/cubilia/curae.aspx

Jacksonville

Florida

Albra

3

☐

Fanoodle

📞(407) 243-0659

https://zimbio.com/omare/consequat/lectus/in/est/rius.js

Orlando

Florida

Albra

4

☐

Wikizz

📞(571) 978-9245

https://zdnet.com/quis/odio/consequat/varius/integer.jion

Fairfax

Virginia

Albra

5

☐

Eidel

📞(585) 311-4895

https://zdnet.com/et/tempus/temper.html

Rochester

New York

Albra

6

☐

Yodo

📞(505) 129-1886

https://yolasite.com/ipsum/aliquam/non/mauris/morbi/non.jpg

Albuquerque

New Mexico

Albra

7

☐

Dubtype

📞(713) 496-7003

https://yelp.com/posuere/metus.xml

Houston

Texas

Albra

8

☐

Rhynoodle

📞(019) 583-4366

https://yellowpages.com/cursus/id/turpis/integer/aliquet/massa.jion

Durham

North Carolina

Albra

9

☐

Yoveo

📞(413) 901-9279

https://yandex.ru/molis.xml

Springfield

Massachusetts

Albra

10

☐

Eidel

📞(202) 539-8823

https://yandex.ru/figula/vehicula/consequat/morbi/a/psum/integer.html

Washington

District of Columbia

Albra

11

☐

Feedspan

📞(786) 629-0261

https://yale.edu/tortor/id/mulla/ultrices/aliquet.aspx

Miami

Florida

Albra

12

☐

Jetpulse

📞(719) 751-8657

https://yale.edu/congue/etiam/justo/etiam/premium/aculis/justo.js

Colorado Springs

Colorado

Albra

13

☐

Miboo

📞(314) 896-4329

https://yahoo.com/magna.aspx

Saint Louis

Missouri

Albra

After this stage, I have set up an Airbyte connection in Salesforce

Connected App Name

airbyte

Help for this Page

Back to List

Custom Apps

Edit

Delete

Manage

Version

1.0

API Name

airbyte

Created Date

11/6/2024, 10:09 AM

Contact Email

By: [Abbas Ibrahim](#)

Contact Phone

ababrahim729@gmail.com

Last Modified Date

11/6/2024, 10:09 AM

By: [Abbas Ibrahim](#)

Description

Info URL

Custom Connected App Handler

Apex Plugin Class

Data Preparation:

In this stage, I've created and configured database objects in Snowflake Datawarehouse so that it can be set up as our destination stage. Raw data will be stored inside the staging table and later transformed to store inside the production table for downstream users.

Below SQL script shows that I've created database in the Snowflake

MSDA3040\_TERMPROJECT.PUBLIC

Settings

1

-- Created a Database for the Term Project

2

CREATE OR REPLACE DATABASE MSDA3040\_TERMPROJECT;

Results

Chart

status

1 Database MSDA3040\_TERMPROJECT successfully created.

Then, created a schema named “MSDA3040\_TERMPROJECT.STAGE\_DATA” within the above Database in order to store the raw data. Below SQL script creates a SQL schema.

```
4 -- Create a Schema within the above DB for the Term Project
5 CREATE OR REPLACE SCHEMA MSDA3040_TERMPROJECT.STAGE_DATA;
```

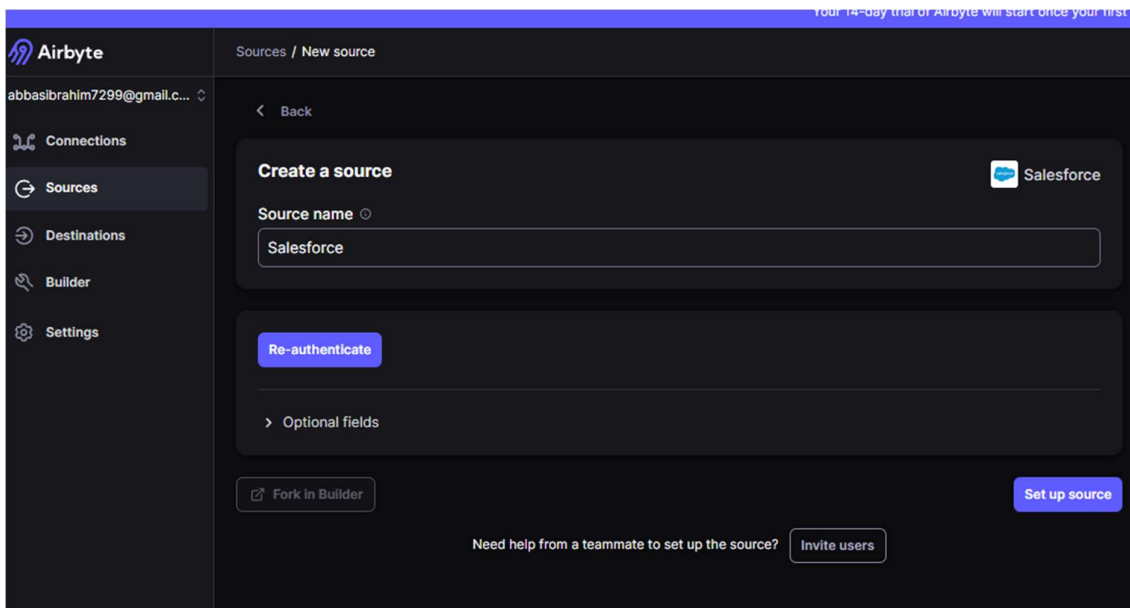
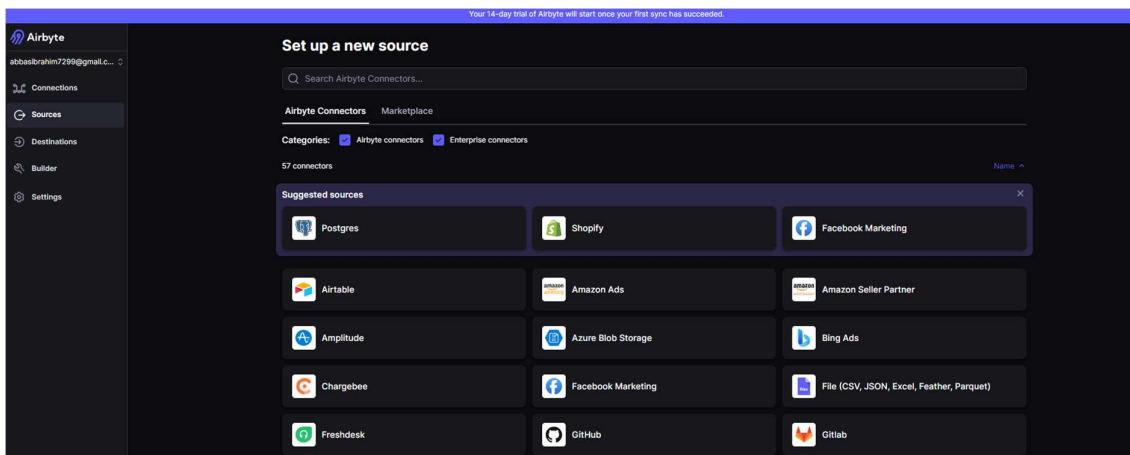
Results Chart

status
1 Schema STAGE_DATA successfully created.

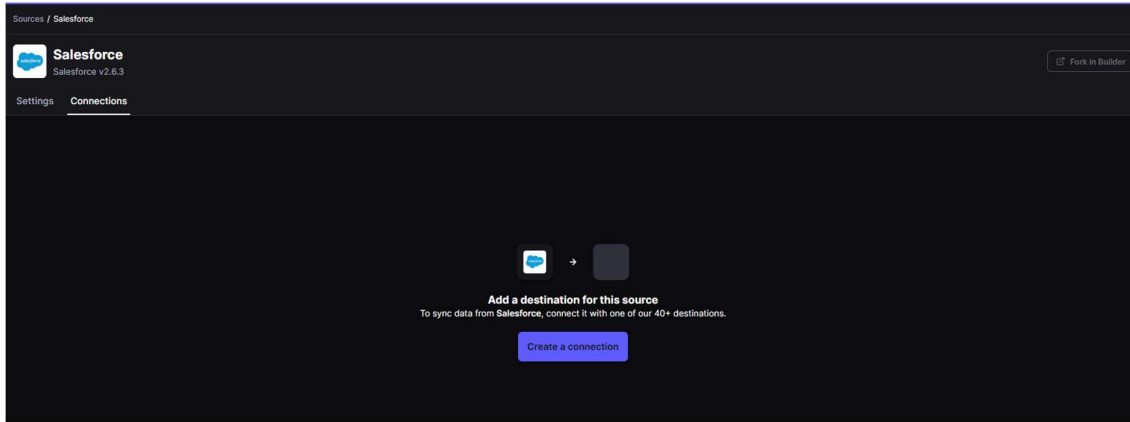
## Data Loading:

Created an Airbyte account to set up the data source and destination to extract and load the data from its source to the destination.

Selected “Salesforce” as ETL tool’s source and have authenticated and configured “Salesforce” as the source for our project.



After successfully setting up the source, added the ETL's destination as Snowflake and filled in the below credentials to accurately load the data to its destination. After filling in all those details, I configured the destination and changed replication frequency to manual and manually selected the fields like ID, Name, Billing City, Billing State, Billing Phone, and Website to load these fields only into the destination data warehouse, also timestamp has been included automatically by the ETL tool itself.



**Create a destination** Snowflake

Destination name

Host   
(account\_name).snowflakecomputing.com or (accountname).(aws\_location).aws.snowflakecomputing.com

Role

Warehouse

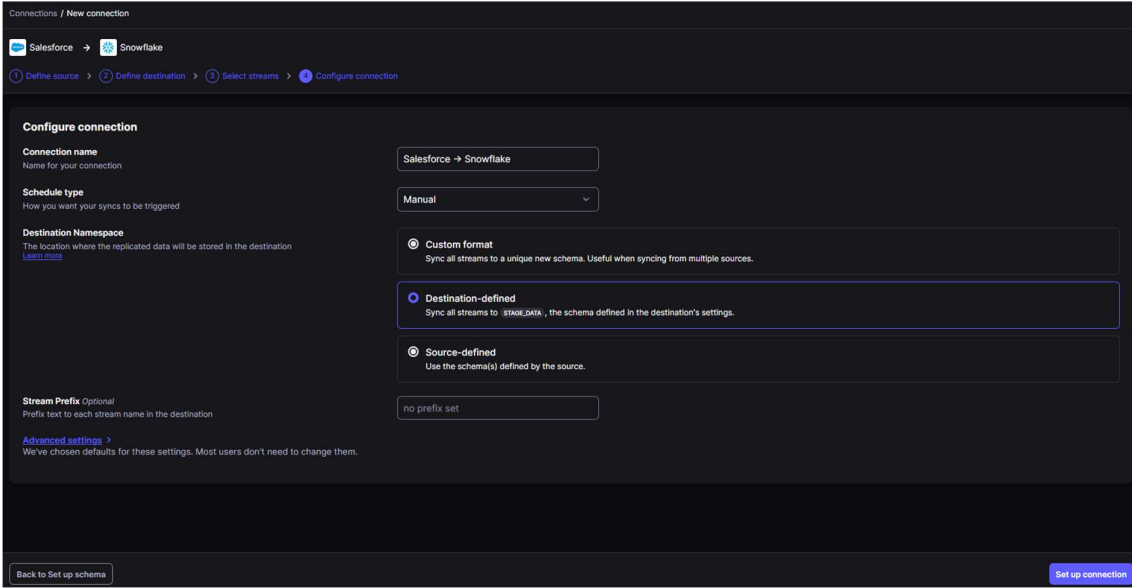
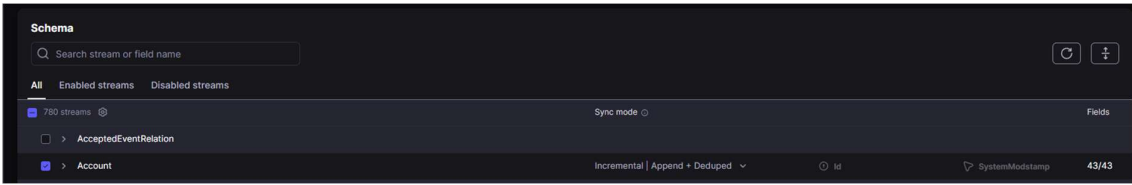
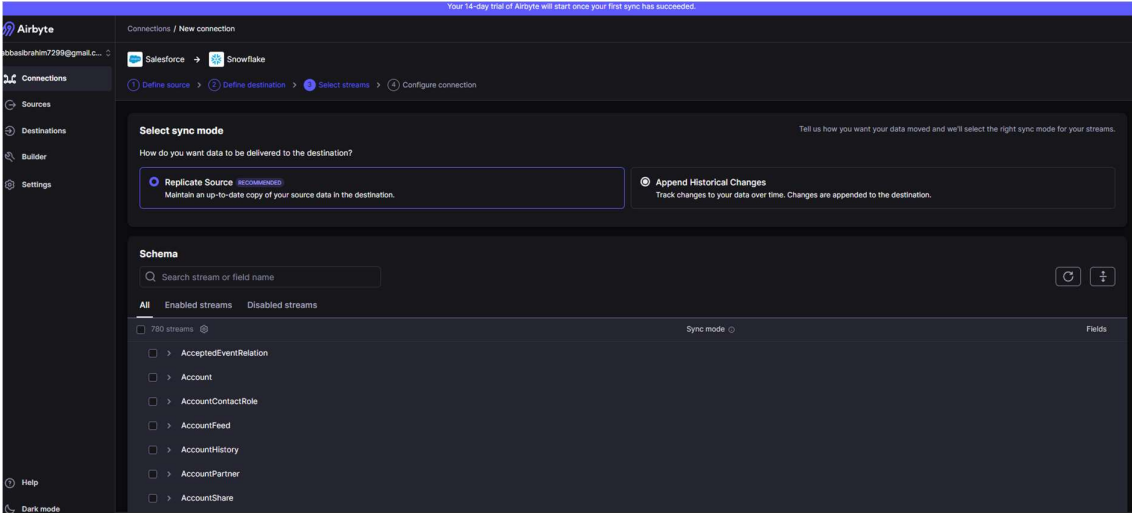
Database

Default Schema

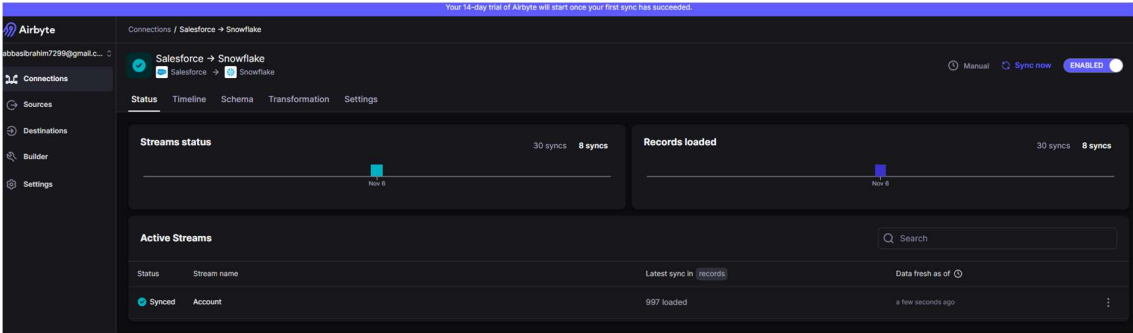
Username

Authorization Method: Username and Password

Password



After all the configuration, click setup connection and sync now to successfully extract the raw data from the salesforce and load the selective fields into the Snowflake data warehouse database object “MSDA3040\_TERMPROJECT.STAGE\_DATA”. Therefore, successfully leveraged airbyte to extract data from Salesforce and loaded it into the staging tables in Snowflake.



Finally, I used the SQL query below to validate whether the Salesforce records were successfully extracted and ingested into the Snowflake by the Airbyte tool for further transformation and data serving.

The screenshot shows a Snowflake SQL query result. The query is: `SELECT * FROM MSDA3040_TERMPROJECT.STAGE_DATA.ACCOUNT;`. The results are displayed in a table with 17 rows. The columns are: `_AIRBYTE_GENERATION_ID`, `ID`, `NAME`, `PHONE`, `WEBSITE`, `BILLINGCITY`, `BILLINGSTATE`, `ACCOUNTSOURCE`, and `SYSTEMMODSTAMP`. The data includes various company names like Dynava, Thoughtsphere, Filpopia, Realcube, Youspan, Fadeo, Browsezoom, Yodel, Brainbox, Snaptags, Edgeblab, Centimia, Edgepulse, Allane, Quinu, Janyx, and Oyope, along with their respective contact information and timestamps.

	_AIRBYTE_GENERATION_ID	ID	NAME	PHONE	WEBSITE	BILLINGCITY	BILLINGSTATE	ACCOUNTSOURCE	SYSTEMMODSTAMP
1	"sync_id": 21455424	0	001Hu00003TFQ83IAH	Dynava	(505) 140-2381	https://howstuffworks	Santa Fe	New Mexico	2024-11-06 18:03:41.000 +0000
2	"sync_id": 21455424	0	001Hu00003TFQDFIAS	Thoughtsphere	(571) 498-9304	https://census.gov/mas	Vienna	Virginia	2024-11-06 18:03:42.000 +0000
3	"sync_id": 21455424	0	001Hu00003TFQFCIAP	Filpopia	(610) 786-7411	https://wp.com/butrice	Philadelphia	Pennsylvania	2024-11-06 18:03:43.000 +0000
4	"sync_id": 21455424	0	001Hu00003TFQEXIAS	Realcube	(466) 851-0530	http://acquirethisname	Garland	Texas	2024-11-06 18:03:43.000 +0000
5	"sync_id": 21455424	0	001Hu00003TFQENIAS	Youspan	(682) 328-7859	http://pagesperso-or	Fort Worth	Texas	2024-11-06 18:03:43.000 +0000
6	"sync_id": 21455424	0	001Hu00003TFQD8IAP	Fadeo	(515) 403-9100	http://mac.com/tellus/	Des Moines	Iowa	2024-11-06 18:03:42.000 +0000
7	"sync_id": 21455424	0	001Hu00003TFQ8yIAH	Browsezoom	(832) 305-1036	https://ifeng.com/odis	Houston	Texas	2024-11-06 18:03:41.000 +0000
8	"sync_id": 21455424	0	001Hu00003TFQAAIAS	Yodel	(260) 247-4500	https://arizona.edu/lee	Fort Wayne	Indiana	2024-11-06 18:03:41.000 +0000
9	"sync_id": 21455424	0	001Hu00003TFQLniIAP	Brainbox	(718) 629-7635	http://list-manage.com	Buffalo	New York	2024-11-06 18:03:45.000 +0000
10	"sync_id": 21455424	0	001Hu00003TFQOqiIAP	Snaptags	(318) 860-4635	http://nh.gov/vin/delfer	Shreveport	Louisiana	2024-11-06 18:03:46.000 +0000
11	"sync_id": 21455424	0	001Hu00003TFQROIAS	Edgeblab	(860) 106-1919	http://stumbleupon.co	Hartford	Connecticut	2024-11-06 18:03:46.000 +0000
12	"sync_id": 21455424	0	001Hu00003TFQ8jIAP	Centimia	(917) 692-0921	http://weba.com/tincid	Brooklyn	New York	2024-11-06 18:03:42.000 +0000
13	"sync_id": 21455424	0	001Hu00003TFQ8diIAP	Edgepulse	(412) 532-2940	http://shop-pro.jp/tincid	Pittsburgh	Pennsylvania	2024-11-06 18:03:46.000 +0000
14	"sync_id": 21455424	0	001Hu00003TFQ8LIAX	Allane	(518) 780-7879	https://harvard.edu/ele	Albany	New York	2024-11-06 18:03:41.000 +0000
15	"sync_id": 21455424	0	001Hu00003TFQ8MIAS	Quinu	(505) 406-9860	http://pbs.org/neque/la	Albuquerque	New Mexico	2024-11-06 18:03:46.000 +0000
16	"sync_id": 21455424	0	001Hu00003TFQOqiIAP	Janyx	(915) 952-8261	http://cisco.com/nunc/	El Paso	Texas	2024-11-06 18:03:46.000 +0000
17	"sync_id": 21455424	0	001Hu00003TFQ8EIAH	Oyope	(704) 702-5101	http://symantec.com/f	Charlotte	North Carolina	2024-11-06 18:03:43.000 +0000