




1. Use **Airbyte** to integrate data sources: customer and related demographic data, web_sales and catalog_sales
 - a. JSON data is converted to columnar data and stored in snowflake




Connections




Sources




Destinations




Builder



Update



Resources



Catalog

Connections / DE-RDS → Snowflake

DE-RDS → Snowflake

Enabled

Status

Job History


Replication

Transformation

Settings

Source						Destination		
Sync	Fields	Namespace	Stream name	Sync mode	Cursor field	Primary key	Namespace	Stream name
<input checked="" type="checkbox"/>	All	raw_st	catalog_sales	Full refresh Append			<destination schema>	catalog_sales
<input checked="" type="checkbox"/>	All	raw_st	customer	Full refresh Overwrite			<destination schema>	customer
<input checked="" type="checkbox"/>	All	raw_st	customer_address	Full refresh Append			<destination schema>	customer_address
<input checked="" type="checkbox"/>	All	raw_st	customer_demographics	Full refresh Append			<destination schema>	customer_demographics
<input checked="" type="checkbox"/>	All	raw_st	date_dim	Full refresh Append			<destination schema>	date_dim
<input checked="" type="checkbox"/>	All	raw_st	household_demographics	Full refresh Append			<destination schema>	household_demographics
<input checked="" type="checkbox"/>	All	raw_st	income_band	Full refresh Append			<destination schema>	income_band
<input checked="" type="checkbox"/>	All	raw_st	item	Full refresh Append			<destination schema>	item
<input checked="" type="checkbox"/>	All	raw_st	promotion	Full refresh Append			<destination schema>	promotion
<input checked="" type="checkbox"/>	All	raw_st	ship_mode	Full refresh Append			<destination schema>	ship_mode
<input checked="" type="checkbox"/>	All	raw_st	time_dim	Full refresh Append			<destination schema>	time_dim
<input checked="" type="checkbox"/>	All	raw_st	warehouse	Full refresh Append			<destination schema>	warehouse
<input checked="" type="checkbox"/>	All	raw_st	web_page	Full refresh Append			<destination schema>	web_page
<input checked="" type="checkbox"/>	All	raw_st	web_sales	Full refresh Append			<destination schema>	web_sales
<input checked="" type="checkbox"/>	All	raw_st	web_site	Full refresh Append			<destination schema>	web_site

2. Airbyte sync runs **daily at 7am UTC** when the hosting **EC2 instance is active**:



Connections / DE-RDS → Snowflake

Connections

Sources

Destinations

Builder

Update

Resources

Settings

0.50.6

DE-RDS → Snowflake

StatusJob HistoryReplicationTransformationSettings

Sync History

Reset your dataSync now

✓ Sync Succeeded	1.21 GB 2,057,234 emitted records 2,057,234 committed records Job id: 19 7m 13s	10:39AM 01/07/2024	⋮
✓ Sync Succeeded	1.2 GB 2,042,056 emitted records 2,042,056 committed records Job id: 18 7m 15s	7:00AM 12/28/2023	⋮
✓ Sync Succeeded	1.2 GB 2,042,056 emitted records 2,042,056 committed records Job id: 17 6m 55s	3:18PM 12/27/2023	⋮
✓ Reset Succeeded (17 streams)	call_center catalog_page catalog_sales customer customer_address customer_demographics date_dim household_demographics ...	3:17PM 12/27/2023	⋮
✓ Sync Succeeded	53.13 MB 100,000 emitted records 100,000 committed records Job id: 15 46s	3:14PM 12/27/2023	⋮
✓ Sync Succeeded	53.13 MB 100,000 emitted records 100,000 committed records Job id: 14 50s	3:09PM 12/27/2023	⋮
✓ Reset Succeeded (1 stream)	customer	3:09PM 12/27/2023	⋮
✓ Sync Succeeded	1.2 GB 2,042,056 emitted records 2,042,056 committed records Job id: 12 6m 50s	2:58PM 12/27/2023	⋮
✓ Sync Succeeded	1.2 GB 2,042,056 emitted records 2,042,056 committed records Job id: 11 7m 15s	7:00AM 12/27/2023	⋮
✓ Sync Succeeded	1.2 GB 2,039,775 emitted records 2,039,775 committed records Job id: 10 6m 55s	7:00AM 12/26/2023	⋮
✓ Sync Succeeded	1.2 GB 2,037,286 emitted records 2,037,286 committed records Job id: 9 6m 32s	7:00AM 12/25/2023	⋮

3. Source tables stored in **snowflake (see below)** used for the data modeling project are from: **AWS lambda function** output (inventory table), and airbyte streams (all other tables)

