



Airbyte y PostgreSQL para ETL con CDC

Rafael Vera Marañón

Contenido

OBJETIVO DEL LABORATORIO	2
• Objetivos específicos	2
DIAGRAMA LÓGICO	2
CAPTURAS DEL PROCESO DE CONFIGURACIÓN	3
• Clonación del repositorio airbyte	3
• Creación de contenedores Docker	4
• Comprobación de contenedores	4
• Inicio de sesión en Airbyte UI	5
• Definición de Fuente de datos para Airbyte	6
• Definición de destino en Airbyte	8
• Definición de conexión en Airbyte	8
CAPTURAS DE OPERACIONES DML Y SINCRONIZACIÓN AIRBYTE	10
• INSERT	10
• UPDATE	11
• DELETE	12
• JOB HISTORY	13
REPOSITORIO	13

OBJETIVO

Implementar una solución de Change Data Capture (CDC) utilizando Airbyte y PostgreSQL, permitiendo la extracción y sincronización de datos modificados en tiempo real.

- Objetivos específicos
 - Desplegar un entorno funcional de Airbyte utilizando Docker.
 - Configurar una base de datos PostgreSQL preparada para CDC.
 - Establecer una conexión en Airbyte que refleje los cambios en las tablas de la base de datos a través de sincronizaciones periódicas.

DIAGRAMA LÓGICO

1. Capa de Ingesta de Datos

- **Airbyte:** Plataforma de integración de datos que se conectará a PostgreSQL para capturar los cambios de datos.

2. Capa de Procesamiento de Streaming

- **Debezium:** Utilizado para la captura de cambios en PostgreSQL y transmitirlos a Airbyte.

3. Capa de Base de Datos y Almacenamiento

- **PostgreSQL:** Base de datos que almacena los datos de empleados y está configurada para CDC.

4. Capa de Visualización

- **Airbyte UI:** Interfaz gráfica utilizada para configurar y monitorizar las conexiones y sincronizaciones.

CAPTURAS DEL PROCESO DE CONFIGURACIÓN

- Clonación del repositorio airbyte

```
git clone --depth=1 https://github.com/airbytehq/airbyte.git
```

```
cd airbyte
```

```
./run-ab-platform.sh
```

```

[airbyte-bottleserver] "logging-call-ast":{"pername:pacakser,gid:127"}
airbyte-bottleserver | 2024-07-21 00:11:19 INFO o.f.d.i.c.n.SLV6[logInfo]:37 - Successfully validated 122 migrations (execution time 00:00:10.83)
airbyte-temporal | ["level":"info","ts":"2024-07-21T00:11:20.120.410Z","msg":"none","service":"matching","component":"matching-engine","ref-task-queue-name":"temporal-sys-per-nos-t","ref-task-queue-type":"Activity","w-f-namespace":"default","lifecycle":"Started","logging-call-ast":{"task-queue-manager-g:330"}]
airbyte-bottleserver | 2024-07-21 00:11:19 WARN i.a.r.v.Loggers$SLV4[Logger(war):299 - [b3d5575c, /127.0.0.1:52437 - R:localhost/127.0.0.1:8125] An exception has been observed post termination, use DEBUG level to see the full stack: java.net.PortUnreachableException: recAddress(...): failed: Connection refused
| 2024-07-21 00:11:19 INFO o.f.d.i.c.n.SLV6[logInfo]:37 - Corrupt version of schema "public": 0.29.0.001
airbyte-bottleserver | 2024-07-21 00:11:19 WARN i.a.r.v.Loggers$SLV4[Logger(war):299 - [b3d5575c, /127.0.0.1:52437 - R:localhost/127.0.0.1:8125] An exception has been observed post termination, use DEBUG level to see the full stack: java.net.PortUnreachableException: recAddress(...): failed: Connection refused
airbyte-bottleserver | 2024-07-21 00:11:19 WARN i.a.r.v.Loggers$SLV4[Logger(war):299 - [b3d5575c, /127.0.0.1:52437 - R:localhost/127.0.0.1:8125] An exception has been observed post termination, use DEBUG level to see the full stack: java.net.PortUnreachableException: recAddress(...): failed: Connection refused
airbyte-bottleserver | 2024-07-21 00:11:19 INFO o.f.d.i.c.n.SLV6[logInfo]:37 - Migrating schema "public" to version "0.30.122.001". Store last sync state
airbyte-bottleserver | 2024-07-21 00:11:19 INFO o.f.d.i.c.n.V0_32_0_001_AirbyteConfigDatabaseDenormalization | Running migration: V0_32_0_001_StoreLastSyncState
airbyte-bottleserver | 2024-07-21 00:11:19 INFO i.a.d.i.c.m.V0_30_32_001_StoreLastSyncState@org.postgresql.jdbc:184 - [Configure migration 0.30.32.001] This is the dev environment; there is no jobs database
airbyte-bottleserver | 2024-07-21 00:11:19 INFO o.f.d.i.c.n.SLV6[logInfo]:37 - Migrating schema "public" to version "0.32.0.001". AirbyteConfigDatabaseDenormalization
airbyte-bottleserver | 2024-07-21 00:11:19 INFO i.a.d.i.c.m.V0_32_0_001_AirbyteConfigDatabaseDenormalization(createNums):82 - source_type enum created
airbyte-bottleserver | 2024-07-21 00:11:19 INFO i.a.d.i.c.m.V0_32_0_001_AirbyteConfigDatabaseDenormalization(createNums):84 - actor_type enum created
airbyte-bottleserver | 2024-07-21 00:11:19 INFO i.a.d.i.c.m.V0_32_0_001_AirbyteConfigDatabaseDenormalization(createNums):86 - operator_type enum created
airbyte-bottleserver | 2024-07-21 00:11:19 INFO i.a.d.i.c.m.V0_32_0_001_AirbyteConfigDatabaseDenormalization(createNums):88 - namespace_definition_type enum created
airbyte-bottleserver | 2024-07-21 00:11:19 INFO i.a.d.i.c.m.V0_32_0_001_AirbyteConfigDatabaseDenormalization(createNums):90 - status_type enum created
airbyte-bottleserver | 2024-07-21 00:11:19 INFO i.a.d.i.c.m.V0_32_0_001_AirbyteConfigDatabaseDenormalization(createAndPopulateWorkspace):112 - workspace_table created
airbyte-bottleserver | 2024-07-21 00:11:19 INFO i.a.d.i.c.m.V0_32_0_001_AirbyteConfigDatabaseDenormalization(createAndPopulateWorkspace):118 - workspace_table populated with 0 records
airbyte-bottleserver | 2024-07-21 00:11:19 INFO i.a.d.i.c.m.V0_32_0_001_AirbyteConfigDatabaseDenormalization(createAndPopulateActorDefinition):191 - actor_definition table created
airbyte-bottleserver | 2024-07-21 00:11:19 INFO i.a.d.i.c.m.V0_32_0_001_AirbyteConfigDatabaseDenormalization(createAndPopulateActorDefinition):191 - actor_definition table populated with 0 source definition records
airbyte-bottleserver | 2024-07-21 00:11:19 INFO i.a.d.i.c.m.V0_32_0_001_AirbyteConfigDatabaseDenormalization(createAndPopulateActorDefinition):230 - actor_definition table populated with 0 destination definition re
records
airbyte-bottleserver | 2024-07-21 00:11:19 INFO i.a.d.i.c.m.V0_32_0_001_AirbyteConfigDatabaseDenormalization(createAndPopulateActor):268 - actor table created
airbyte-bottleserver | 2024-07-21 00:11:19 INFO i.a.d.i.c.m.V0_32_0_001_AirbyteConfigDatabaseDenormalization(createAndPopulateActor):303 - actor table populated with 0 source records
airbyte-bottleserver | 2024-07-21 00:11:19 INFO i.a.d.i.c.m.V0_32_0_001_AirbyteConfigDatabaseDenormalization(createAndPopulateActor):339 - actor table populated with 0 destination records
airbyte-bottleserver | 2024-07-21 00:11:19 INFO i.a.d.i.c.m.V0_32_0_001_AirbyteConfigDatabaseDenormalization(createAndPopulateActorAuthParameter):406 - actor_auth_parameter table created
airbyte-bottleserver | 2024-07-21 00:11:19 INFO i.a.d.i.c.m.V0_32_0_001_AirbyteConfigDatabaseDenormalization(createAndPopulateActorAuthParameter):440 - actor_auth_parameter table populated with 0 source auth para
as records
airbyte-bottleserver | 2024-07-21 00:11:19 INFO i.a.d.i.c.m.V0_32_0_001_AirbyteConfigDatabaseDenormalization(createAndPopulateActorAuthParameter):475 - actor_auth_parameter table populated with 0 destination auth
params records
airbyte-bottleserver | 2024-07-21 00:11:19 INFO i.a.d.i.c.m.V0_32_0_001_AirbyteConfigDatabaseDenormalization(createAndPopulateOperation):505 - operation table created
airbyte-bottleserver | 2024-07-21 00:11:19 INFO i.a.d.i.c.m.V0_32_0_001_AirbyteConfigDatabaseDenormalization(createAndPopulateOperation):534 - operation table populated with 0 records

```

- Creación de contenedores Docker

Containers [Give feedback](#)

Container CPU usage 3.96% / 700% (7 CPUs available) Container memory usage 2.52GB / 5.66GB [Show charts](#)

Q Search Only show running containers

<input type="checkbox"/>	Name	Image	Status	CPU (%)	Port(s)	Last started	Actions
<input type="checkbox"/>	airbyte_cdc_project		Running (1/1)	0%		23 minutes ago	
<input type="checkbox"/>	postgres-1 369e04625a3f	debezium/postgres:13	Running	0%	5432-5432	23 minutes ago	
<input type="checkbox"/>	airbyte		Running (10/1)	3.96%		27 minutes ago	
<input type="checkbox"/>	init 1ec48215b10e	airbyte/init:0.63.8	Exited	0%		28 minutes ago	
<input type="checkbox"/>	temporal 410e43a6a7a0	airbyte/temporal:0.63.8	Running	2.19%		28 minutes ago	
<input type="checkbox"/>	docker-proxy-1 8afd72f12f6e	alpine/socat	Running	0%	53628-2375	28 minutes ago	
<input type="checkbox"/>	db aed8577c12dc	airbyte/db:0.63.8	Running	0.34%		28 minutes ago	
<input type="checkbox"/>	bootloader d7fc30d3e989	airbyte/bootloader:0.63.8	Exited	0%		28 minutes ago	
<input type="checkbox"/>	webapp 9708054f8c3c	airbyte/webapp:0.63.8	Running	0%		27 minutes ago	
<input type="checkbox"/>	worker 9c6aaddc9cdf	airbyte/worker:0.63.8	Running	0.37%		27 minutes ago	
<input type="checkbox"/>	connector-builder-server 2fdb9a6539f	airbyte/connector-builder-server:0.63.8	Running	0.09%		27 minutes ago	
<input type="checkbox"/>	api-server 539b1b3f5dc1	airbyte/airbyte-api-server:0.63.8	Running	0.14%		27 minutes ago	

- Comprobación de contenedores

```
(venv) PS C:\Users\jriver\PycharmProjects\Python\airbyte_cdc_project> docker ps
```

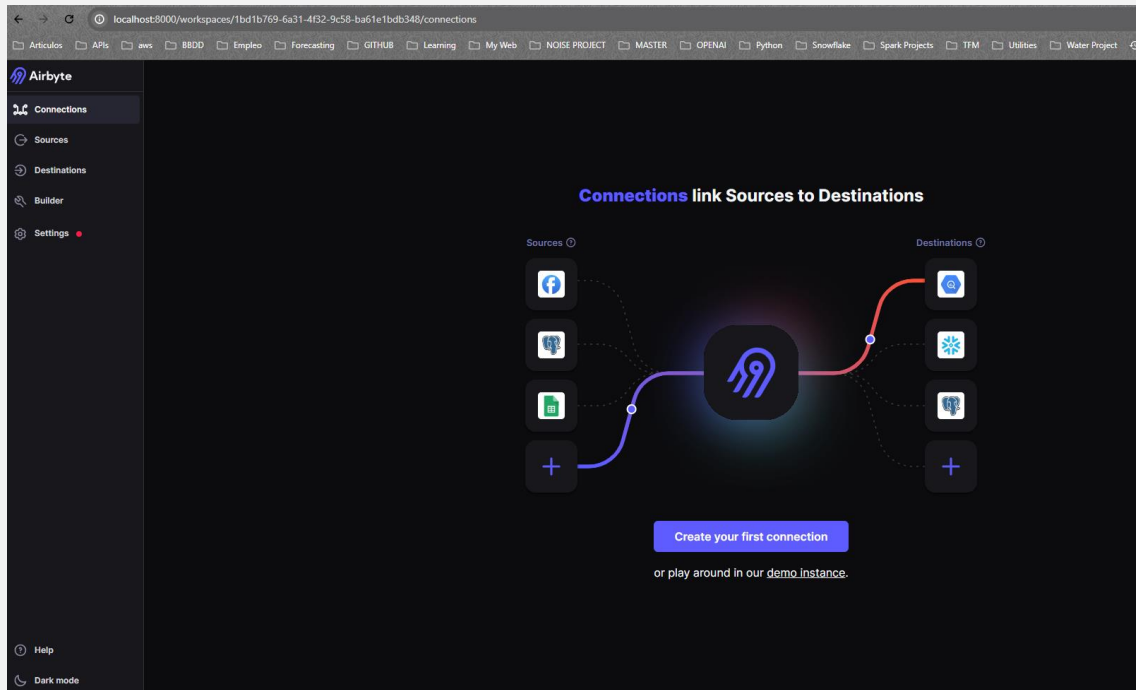
CONTAINER ID	IMAGE	COMMAND	CREATED	STATUS	PORTS	NAMES
62272448f0db	airbyte/proxy:0.63.8	"/run.sh ./run.sh"	28 minutes ago	Up 28 minutes	0.0.0.0:8080-8081->8080-8081/tcp, 0.0.0.0:8083->8083/tcp, 80/tcp, 0.0.0.0:8086->8086/tcp	airbyte-proxy
a735f96db0a8	airbyte/server:0.63.8	"/bin/bash -c airbyt"	28 minutes ago	Up 28 minutes	5005/tcp, 8080/tcp, 0.0.0.0:54338->8081/tcp	airbyte-server
9c0a8dc9cdf	airbyte/worker:0.63.8	"/bin/bash -c airbyt"	28 minutes ago	Up 28 minutes	5005/tcp, 0.0.0.0:54328->9080/tcp	airbyte-worker
499acd08a976	airbyte/cron:0.63.8	"/bin/bash -c airbyt"	28 minutes ago	Up 28 minutes		airbyte-cron
539b1b3f5dc1	airbyte/airbyte-api-server:0.63.8	"/bin/bash -c airbyt"	28 minutes ago	Up 28 minutes	5005/tcp, 0.0.0.0:54332->8086/tcp	airbyte-api-server
2fdb9a6539f	airbyte/connector-builder-server:0.63.8	"/bin/bash -c airbyt"	28 minutes ago	Up 28 minutes	0.0.0.0:54331->8080/tcp	airbyte-connector-b
builder-server						
9708054f8c3c	airbyte/webapp:0.63.8	"/docker-entrypoint.s"	28 minutes ago	Up 28 minutes	8080/tcp	airbyte-webapp
410e43a6a7a0	airbyte/temporal:0.63.8	"/docker-entrypoint.s"	28 minutes ago	Up 28 minutes	6933-6925/tcp, 6934/tcp, 7233-7235/tcp, 7239/tcp	airbyte-temporal
8afd72f12f6e	alpine/socat	"socat -t 900 TCP-LI"	28 minutes ago	Up 28 minutes	0.0.0.0:53628->2375/tcp	airbyte-docker-prox
y-1						
aed8577c12dc	airbyte/db:0.63.8	"/docker-entrypoint.s"	28 minutes ago	Up 28 minutes	5432/tcp	airbyte-db
369e04625a3f	debezium/postgres:13	"/docker-entrypoint.s"	About an hour ago	Up 23 minutes	0.0.0.0:5432->5432/tcp	airbyte_cdc_project
-postgres-1						

```
(venv) PS C:\Users\jriver\PycharmProjects\Python\airbyte_cdc_project>
```

- Inicio de sesión en Airbyte UI

Usuario: airbyte

Contraseña: password



- Definición de Fuente de datos para Airbyte

El host para ser reconocido correctamente tuve que indicar la ip del contenedor postgres en Docker.

The screenshot shows the Airbyte web interface in a browser window. The address bar indicates the URL: `localhost:8000/workspaces/1bd1b769-6a31-4f32-9c58-ba61e1bdb348/source/e6da2146-122a-4317-8bb7-9fb0a052503e`. The browser's tab bar shows several open tabs: Articulos, APIs, aws, BBDD, Empleo, Forecasting, GITHUB, Learning, My Web, NOISE PROJECT, and MASTER. The Airbyte interface has a dark theme. On the left is a sidebar with navigation links: Connections, Sources (selected), Destinations, Builder, and Settings. The main content area is titled 'Sources / Postgres' and shows the configuration for a 'Postgres' source (version v3.4.24). There are two tabs: 'Settings' (active) and 'Connections'. The 'Settings' tab contains several sections: 'Source Settings' with fields for 'Source name' (Postgres), 'Host' (172.18.0.3), 'Port' (5432), and 'Database Name' (airbyte); a section for 'Optional fields'; 'Username' (airbyte) and 'Password' (masked with asterisks, with an 'Edit' button); and a 'Security' section with 'SSL Modes' set to 'disable' and 'SSH Tunnel Method' set to 'No Tunnel'.

localhost:8000/workspaces/1bd1b769-6a31-4f32-9c58-ba61e1bdb348/source/e6da2146-122a-4317-8bb7-9fb0a052503e

awsBBDDEmpleoForecastingGITHUBLearningMy WebNOISE PROJECTMASTER

Advanced

Update Method ⓘ

☒ Read Changes using Write-Ahead Log (CDC)
Recommended - Incrementally reads new inserts, updates, and deletes using the Postgres [write-ahead log \(WAL\)](#). This needs to be configured on the source database itself. Recommended for tables of any size.

☐ Detect Changes with Xmin System Column
Recommended - Incrementally reads new inserts and updates via Postgres [Xmin system column](#). Suitable for databases that have low transaction pressure.

☐ Scan Changes with User Defined Cursor
Incrementally detects new inserts and updates using the [cursor column](#) chosen when configuring a connection (e.g. created_at, updated_at).

Replication Slot ⓘ

airbyte_slot

Publication ⓘ

my_publication

Optional fields

Initial Waiting Time in Seconds (Advanced) ⓘ Optional

2400

Size of the queue (Advanced) ⓘ Optional

10000

LSN commit behaviour ⓘ Optional

After loading Data in the destination

Debezium heartbeat query (Advanced) ⓘ Optional

Invalid CDC position behavior (Advanced) ⓘ Optional

Fail sync

Optional fields

Test the source

Retest saved source

Delete this source

Cancel

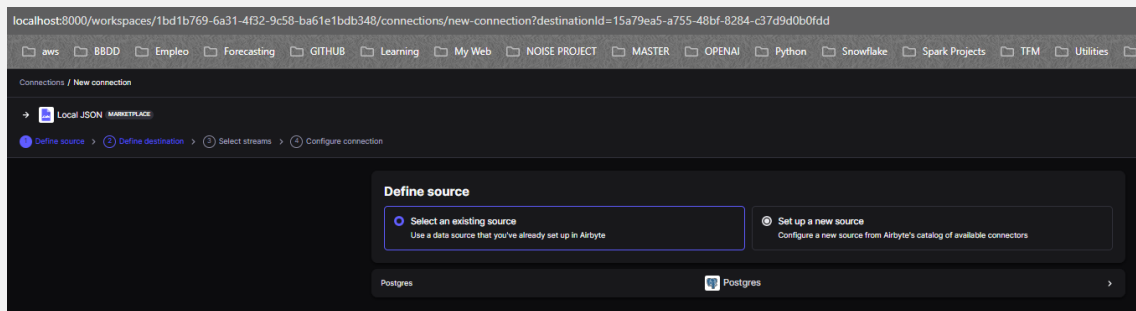
Test and save

7

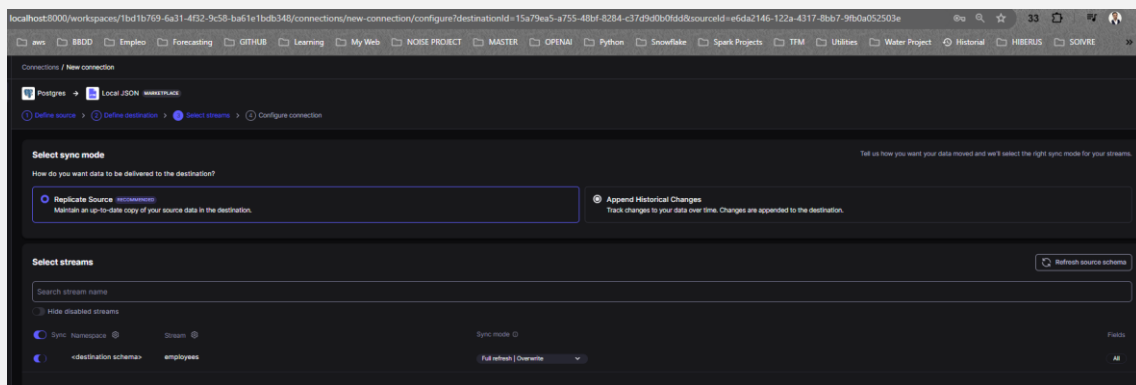
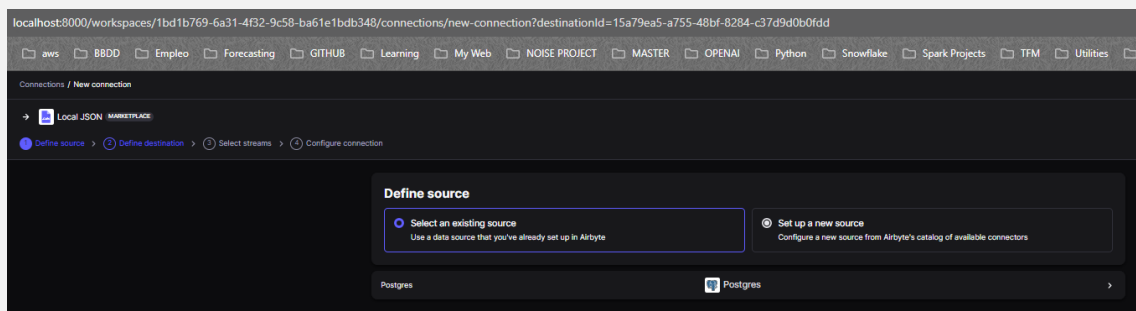
Rafael Vera Marañón

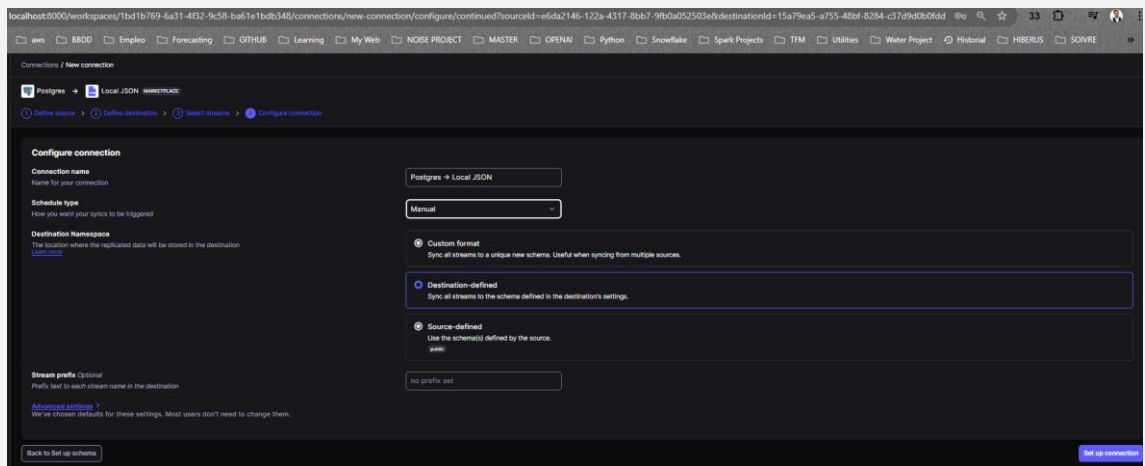
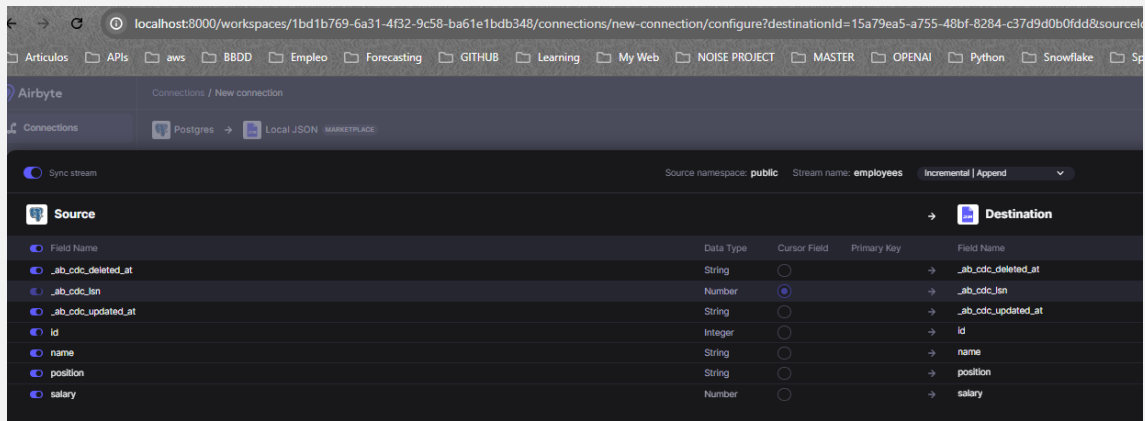
- Definición de destino en Airbyte

Se define un destino Local Json para el directorio local/airbyte_output



- Definición de conexión en Airbyte





CAPTURAS DE OPERACIONES DML Y SINCRONIZACIÓN AIRBYTE

- INSERT

En primer lugar, se muestra el acceso al contenedor postgres en dokcer y consulta la tabla existente sin registros.

Posteriormente se ejecuta sql para insert y se muestra tabla actualizada.

```
Is the server running on that host and accepting TCP/IP connections?
root@airbyte-abctl-control-plane:/#

What's next:
Try Docker Debug for seamless, persistent debugging tools in any container or image → docker debug 3d9831e83b01
Learn more at https://docs.docker.com/go/debug-cli/

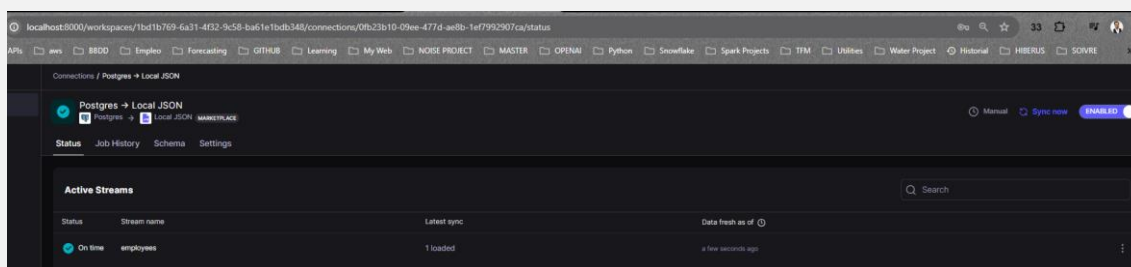
(venv) PS C:\Users\jvrer\PycharmProjects\Python\airbyte_cdc_project> docker ps
CONTAINER ID   IMAGE                                COMMAND                  CREATED        STATUS        PORTS                               NAMES
fe5ce4439cb2   debezium/postgres:13                "docker-entrypoint.s..." 38 minutes ago Up 38 minutes 0.0.0.0:5432->5432/tcp, :::5432->5432/tcp   airbyte_cdc_project-postgres-1
3d9831e83b01   kindest/node:v1.29.4               "/usr/local/bin/entr..." 47 minutes ago Up 47 minutes 0.0.0.0:8000->80/tcp, 127.0.0.1:52631->6443/tcp   airbyte-abctl-control-plane

(venv) PS C:\Users\jvrer\PycharmProjects\Python\airbyte_cdc_project> docker exec -it fe5ce4439cb2 /bin/bash
root@fe5ce4439cb2:/# psql -U airbyte -d airbyte -c "\dt"
List of relations
Schema | Name  | Type  | Owner
-----+-----+-----+-----
public | employees | table | airbyte
(1 row)

root@fe5ce4439cb2:/# psql -U airbyte -d airbyte -c "SELECT * FROM employees;"
id | name | position | salary
---+---+---+---
(0 rows)

root@fe5ce4439cb2:/# psql -U airbyte -d airbyte -c "INSERT INTO employees (name, position, salary) VALUES ('John Doe', 'Developer', 75000);"
INSERT 0 1
root@fe5ce4439cb2:/# psql -U airbyte -d airbyte -c "SELECT * FROM employees;"
id | name | position | salary
---+---+---+---
1 | John Doe | Developer | 75000
(1 row)
```

Manualmente se lanza una sincronización con resultado exitoso desde el UI de Airbyte



A continuación, se accede al contenedor airbyte-server en Docker para recuperar y mostrar el archivo json donde persiste la información de los cambios en postgres capturada por airbyte.

```
(venv) PS C:\Users\jvrer\PycharmProjects\Python\airbyte_cdc_project> docker exec -it airbyte-server /bin/bash
airbyte@ef35f90ba08 app$ cd /tmp/airbyte.local/airbyte_output
airbyte@ef35f90ba08 airbyte_output$ ls
airbyte_raw_employees.jsonl
airbyte@ef35f90ba08 airbyte_output$ cat _airbyte_raw_employees.jsonl
{"_airbyte_ab_id":"d8dc9cc-582c-4749-8d70-54b0e52a1a3","_airbyte_emitted_at":"1721321365025","_airbyte_data":{"id":1,"name":"John Doe","position":"Developer","salary":75000,"_ab_cdc_updated_at":"2024-07-21T00:22:45.825488544Z"},"_ab_cdc_idm":"23938840","_ab_cdc_deleted_at":null}
airbyte@ef35f90ba08 airbyte_output$
```

Se ejecuta un insert adicional:

Instale la versión más reciente de PowerShell para obtener nuevas características y mejoras. <https://aka.ms/PSWindows>

```
(venv) PS C:\Users\jrver\PcharmProjects\Python\airbyte_cdc_project> docker exec -it 369e04625a3f /bin/bash
root@369e04625a3f:/# psql -U airbyte -d airbyte
psql (13.15 (Debian 13.15-1.pgdg110+1))
Type "help" for help.

airbyte=# INSERT INTO employees (name, position, salary) VALUES ('Jane Smith', 'Data Scientist', 80000);
INSERT 0 1
airbyte=#
```

Y se recupera el json actualizado:

```
(venv) PS C:\Users\jrver\PcharmProjects\Python\airbyte_cdc_project> docker exec -it airbyte-server /bin/bash
airbyte@ef35f90ba08 app]$ cd /tmp/airbyte_local/airbyte_output
airbyte@ef35f90ba08 airbyte_output]$ ls
airbyte_raw_employees.jsonl
airbyte@ef35f90ba08 airbyte_output]$ cat airbyte_raw_employees.jsonl
{"_airbyte_ab_id":"d8dcf9cc-582c-4749-8d70-540b3e2a1a3","_airbyte_emitted_at":"1721521365025","_airbyte_data":{"id":1,"name":"John Doe","position":"Developer","salary":75000,"_ab_cdc_updated_at":"2024-07-21T00:22:45.025480344Z"},"_ab_cdc_lsn":"23938560","_ab_cdc_deleted_at":null}
airbyte@ef35f90ba08 airbyte_output]$ cat airbyte_raw_employees.jsonl
{"_airbyte_ab_id":"d8dcf9cc-582c-4749-8d70-540b3e2a1a3","_airbyte_emitted_at":"1721521365025","_airbyte_data":{"id":1,"name":"John Doe","position":"Developer","salary":75000,"_ab_cdc_updated_at":"2024-07-21T00:22:45.025480344Z"},"_ab_cdc_lsn":"23938560","_ab_cdc_deleted_at":null}
{"_airbyte_ab_id":"f26e2f1-fd14-4916-bd18-e2c760bb43d","_airbyte_emitted_at":"1721522981099","_airbyte_data":{"id":2,"name":"Jane Smith","position":"Data Scientist","salary":80000,"_ab_cdc_updated_at":"2024-07-21T00:48:49.2192"},"_ab_cdc_lsn":"23985272","_ab_cdc_deleted_at":null}
airbyte@ef35f90ba08 airbyte_output]$
```

- UPDATE

```
(venv) PS C:\Users\jrver\PcharmProjects\Python\airbyte_cdc_project> docker exec -it 369e04625a3f /bin/bash
root@369e04625a3f:/# psql -U airbyte -d airbyte
psql (13.15 (Debian 13.15-1.pgdg110+1))
Type "help" for help.
```

```
airbyte=# INSERT INTO employees (name, position, salary) VALUES ('Jane Smith', 'Data Scientist', 80000);
INSERT 0 1
airbyte=# SELECT * FROM employees;
 id | name | position | salary
-----+-----+-----+-----
  1 | John Doe | Developer | 75000
  2 | Jane Smith | Data Scientist | 80000
(2 rows)
```

```
airbyte=# UPDATE employees SET position = 'Data Engineer' WHERE name = 'John Doe';
UPDATE 1
airbyte=# SELECT * FROM employees;
 id | name | position | salary
-----+-----+-----+-----
  2 | Jane Smith | Data Scientist | 80000
  1 | John Doe | Data Engineer | 75000
(2 rows)
```

```
airbyte=#
```

```
(venv) PS C:\Users\jrver\PcharmProjects\Python\airbyte_cdc_project> docker exec -it airbyte-server /bin/bash
airbyte@ef35f90ba08 app]$ cd /tmp/airbyte_local/airbyte_output
airbyte@ef35f90ba08 airbyte_output]$ ls
airbyte_raw_employees.jsonl
airbyte@ef35f90ba08 airbyte_output]$ cat airbyte_raw_employees.jsonl
{"_airbyte_ab_id":"d8dcf9cc-582c-4749-8d70-540b3e2a1a3","_airbyte_emitted_at":"1721521365025","_airbyte_data":{"id":1,"name":"John Doe","position":"Developer","salary":75000,"_ab_cdc_updated_at":"2024-07-21T00:22:45.025480344Z"},"_ab_cdc_lsn":"23938560","_ab_cdc_deleted_at":null}
airbyte@ef35f90ba08 airbyte_output]$ cat airbyte_raw_employees.jsonl
{"_airbyte_ab_id":"d8dcf9cc-582c-4749-8d70-540b3e2a1a3","_airbyte_emitted_at":"1721521365025","_airbyte_data":{"id":1,"name":"John Doe","position":"Developer","salary":75000,"_ab_cdc_updated_at":"2024-07-21T00:22:45.025480344Z"},"_ab_cdc_lsn":"23938560","_ab_cdc_deleted_at":null}
{"_airbyte_ab_id":"f26e2f1-fd14-4916-bd18-e2c760bb43d","_airbyte_emitted_at":"1721522981099","_airbyte_data":{"id":2,"name":"Jane Smith","position":"Data Scientist","salary":80000,"_ab_cdc_updated_at":"2024-07-21T00:48:49.2192"},"_ab_cdc_lsn":"23985272","_ab_cdc_deleted_at":null}
airbyte@ef35f90ba08 airbyte_output]$ cat airbyte_raw_employees.jsonl
{"_airbyte_ab_id":"f26e2f1-fd14-4916-bd18-e2c760bb43d","_airbyte_emitted_at":"1721522981099","_airbyte_data":{"id":2,"name":"Jane Smith","position":"Data Scientist","salary":80000,"_ab_cdc_updated_at":"2024-07-21T00:48:49.2192"},"_ab_cdc_lsn":"23985272","_ab_cdc_deleted_at":null}
{"_airbyte_ab_id":"1a2941b7-5e1b-48ad-ad05-e98029c4000","_airbyte_emitted_at":"17215232035821","_airbyte_data":{"id":1,"name":"John Doe","position":"Data Engineer","salary":75000,"_ab_cdc_updated_at":"2024-07-21T00:52:44.651Z"},"_ab_cdc_lsn":"24039848","_ab_cdc_deleted_at":null}
airbyte@ef35f90ba08 airbyte_output]$
```

- DELETE

```
(venv) PS C:\Users\jrver\PycharmProjects\Python\airbyte_cdc_project> docker exec -it 369e04625a3f /bin/bash
root@369e04625a3f:/# psql -U airbyte -d airbyte
psql (13.15 (Debian 13.15-1.pgdg110+1))
Type "help" for help.

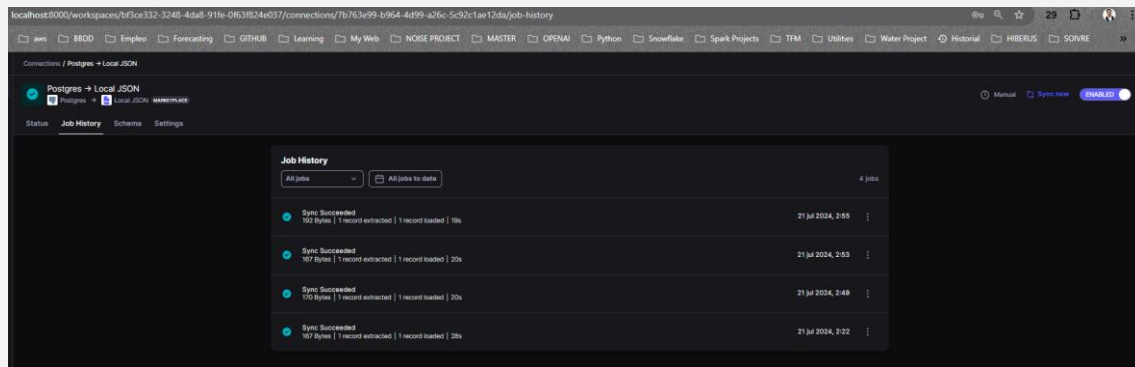
airbyte=# INSERT INTO employees (name, position, salary) VALUES ('Jane Smith', 'Data Scientist', 80000);
INSERT 0 1
airbyte=# SELECT * FROM employees;
 id |   name   |   position   | salary 
----+-----+-----+-----
  1 | John Doe | Developer    |  75000
  2 | Jane Smith | Data Scientist |  80000
(2 rows)

airbyte=# UPDATE employees SET position = 'Data Engineer' WHERE name = 'John Doe';
UPDATE 1
airbyte=# SELECT * FROM employees;
 id |   name   |   position   | salary 
----+-----+-----+-----
  2 | Jane Smith | Data Scientist |  80000
  1 | John Doe  | Data Engineer |  75000
(2 rows)

airbyte=# DELETE FROM employees WHERE name = 'Jane Smith';
DELETE 1
airbyte=# SELECT * FROM employees;
 id |   name   |   position   | salary 
----+-----+-----+-----
  1 | John Doe | Data Engineer |  75000
```

```
(venv) PS C:\Users\jrver\PycharmProjects\Python\airbyte_cdc_project> docker exec -i airbyte-server /bin/bash
[airbyte@ef35f90ba08 app]$ cd /tmp/airbyte_local/airbyte_output
[airbyte@ef35f90ba08 airbyte_output]$ ls
airbyte_raw_employees.jsonl
[airbyte@ef35f90ba08 airbyte_output]$ cat _airbyte_raw_employees.jsonl
{"_airbyte_ab_id":"dbcf9cc-582c-4749-8d70-540b3e52a1a3","_airbyte_emitted_at":1721521365025,"_airbyte_data":{"id":1,"name":"John Doe","position":"Developer","salary":75000,"_ab_cdc_updated_at":"2024-07-21T00:22:45.025480346Z"},"_ab_cdc_lsn":23938560,"_ab_cdc_deleted_at":null}
[airbyte@ef35f90ba08 airbyte_output]$ cat _airbyte_raw_employees.jsonl
{"_airbyte_ab_id":"dbcf9cc-582c-4749-8d70-540b3e52a1a3","_airbyte_emitted_at":1721521365025,"_airbyte_data":{"id":1,"name":"John Doe","position":"Developer","salary":75000,"_ab_cdc_updated_at":"2024-07-21T00:22:45.025480346Z"},"_ab_cdc_lsn":23938560,"_ab_cdc_deleted_at":null}
{"_airbyte_ab_id":"f26e43f1-fd16-4916-bd18-d2c70b0b43d","_airbyte_emitted_at":1721522981099,"_airbyte_data":{"id":2,"name":"Jane Smith","position":"Data Scientist","salary":80000,"_ab_cdc_updated_at":"2024-07-21T00:48:49.219578312Z"},"_ab_cdc_lsn":23938560,"_ab_cdc_deleted_at":null}
[airbyte@ef35f90ba08 airbyte_output]$ cat _airbyte_raw_employees.jsonl
{"_airbyte_ab_id":"f26e43f1-fd16-4916-bd18-d2c70b0b43d","_airbyte_emitted_at":1721522981099,"_airbyte_data":{"id":2,"name":"Jane Smith","position":"Data Scientist","salary":80000,"_ab_cdc_updated_at":"2024-07-21T00:48:49.219578312Z"},"_ab_cdc_lsn":23938560,"_ab_cdc_deleted_at":null}
{"_airbyte_ab_id":"1a2941b7-5e1b-48ad-adb5-998d029c0400","_airbyte_emitted_at":1721523203821,"_airbyte_data":{"id":1,"name":"John Doe","position":"Data Engineer","salary":75000,"_ab_cdc_updated_at":"2024-07-21T00:52:44.651278312Z"},"_ab_cdc_lsn":24039848,"_ab_cdc_deleted_at":null}
[airbyte@ef35f90ba08 airbyte_output]$ cat _airbyte_raw_employees.jsonl
{"_airbyte_ab_id":"dbcf9cc-582c-4749-8d70-540b3e52a1a3","_airbyte_emitted_at":1721521365025,"_airbyte_data":{"id":1,"name":"John Doe","position":"Developer","salary":75000,"_ab_cdc_updated_at":"2024-07-21T00:22:45.025480346Z"},"_ab_cdc_lsn":23938560,"_ab_cdc_deleted_at":null}
{"_airbyte_ab_id":"f26e43f1-fd16-4916-bd18-d2c70b0b43d","_airbyte_emitted_at":1721522981099,"_airbyte_data":{"id":2,"name":"Jane Smith","position":"Data Scientist","salary":80000,"_ab_cdc_updated_at":"2024-07-21T00:48:49.219578312Z"},"_ab_cdc_lsn":23938560,"_ab_cdc_deleted_at":null}
{"_airbyte_ab_id":"1a2941b7-5e1b-48ad-adb5-998d029c0400","_airbyte_emitted_at":1721523203821,"_airbyte_data":{"id":1,"name":"John Doe","position":"Data Engineer","salary":75000,"_ab_cdc_updated_at":"2024-07-21T00:52:44.651278312Z"},"_ab_cdc_lsn":24039848,"_ab_cdc_deleted_at":null}
{"_airbyte_ab_id":"d74c52bf-a85e-49df-84dd-9d19e35da63","_airbyte_emitted_at":1721523332357,"_airbyte_data":{"id":2,"name":"Jane Smith","position":"Data Scientist","salary":80000,"_ab_cdc_updated_at":"2024-07-21T00:54:49.915278312Z"},"_ab_cdc_lsn":24087120,"_ab_cdc_deleted_at":"2024-07-21T00:54:49.915278312Z"}
[airbyte@ef35f90ba08 airbyte_output]$
```

- JOB HISTORY



REPOSITORIO

https://github.com/Rafavermar/airbyte_cdc_etl_project