

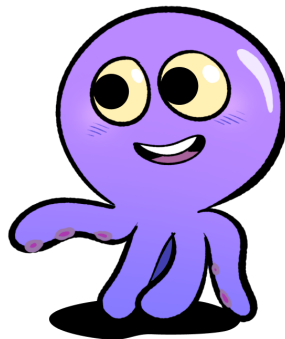


Trabajo Práctico

Laboratorio de Sistemas Operativos y Redes

Primer cuatrimestre - 2023

Deploy de Airbyte



Alan Nahuel Pacheco

Índice

Objetivo del trabajo práctico	3
Entorno de trabajo	3
¿Qué es Airbyte?	3
Instalación	4
Casos de uso	6
Contribución	6
Laboratorio	6
Referencias	8

Objetivo del trabajo práctico

El presente trabajo práctico tiene como objetivo explicar las características de la herramienta Airbyte, requisitos de instalación, su despliegue y sus respectivos casos de uso. También se presentan diferentes escenarios para su despliegue.

Entorno de trabajo

La plataforma se desplegó en una máquina virtual con Debian 12.

Las características son las siguientes:

Procesador: AMD Ryzen 5 5600X

Almacenamiento: 50 gb

Memoria RAM: 4gb

¿Qué es Airbyte?

Es una herramienta de código abierto que nos permite extraer y cargar datos de diversas fuentes, ya sea desde una API, otra base de datos o aplicaciones, hacia nuestras bases de datos de big data. Esta plataforma nos brinda la posibilidad de centralizar nuestros datos en un único lugar, permitiéndonos relacionarlos de manera eficiente.

Instalación

La ventaja que tiene esta herramienta es la fácil instalación y despliegue. Previamente, hay que tener instalado docker y docker compose, cuyos pasos de instalación se detallan en los siguientes links:

- Docker: <https://docs.docker.com/engine/install/>
- Docker Compose: <https://docs.docker.com/compose/install/>

Una vez cumplidos los prerequisites, para hacer el despliegue de la herramienta hay que hacer los siguientes pasos:

1. Clonar el repositorio
2. Dirigirnos a la carpeta del proyecto
3. Ejecutar el script de instalación

```
$ git clone https://github.com/airbytehq/airbyte.git
$ cd airbyte
$ ./run-ab-platform.sh
```

Dentro del script, descarga los archivos necesarios (docker-compose, flags.yml, .env) y ejecuta el archivo yaml del docker-compose

Siguiendo estos pasos, debería aparecer lo siguiente, y posteriormente empezar a crear los servicios con docker compose

```
nahuel@nahuel:~$ git clone https://github.com/airbytehq/airbyte.git
Cloning into 'airbyte'...
remote: Enumerating objects: 2240355, done.
remote: Counting objects: 100% (55138/55138), done.
remote: Compressing objects: 100% (2576/2576), done.
remote: Total 2240355 (delta 29193), reused 53952 (delta 28223), pack-reused 2185217
Receiving objects: 100% (2240355/2240355), 3.34 GiB | 34.06 MiB/s, done.
Resolving deltas: 100% (1211691/1211691), done.
Updating files: 100% (15342/15342), done.
nahuel@nahuel:~$ cd airbyte/
nahuel@nahuel:~/airbyte$ ./run-ab-platform.sh

AIRBYTE
Move Data

Downloading docker-compose.yaml
Downloading docker-compose.debug.yaml
Downloading .env
Downloading .env.dev
Downloading flags.yml
Loading Shell Variables from .env...
Loading Shell Variables from .env.dev...

Starting Docker Compose
```

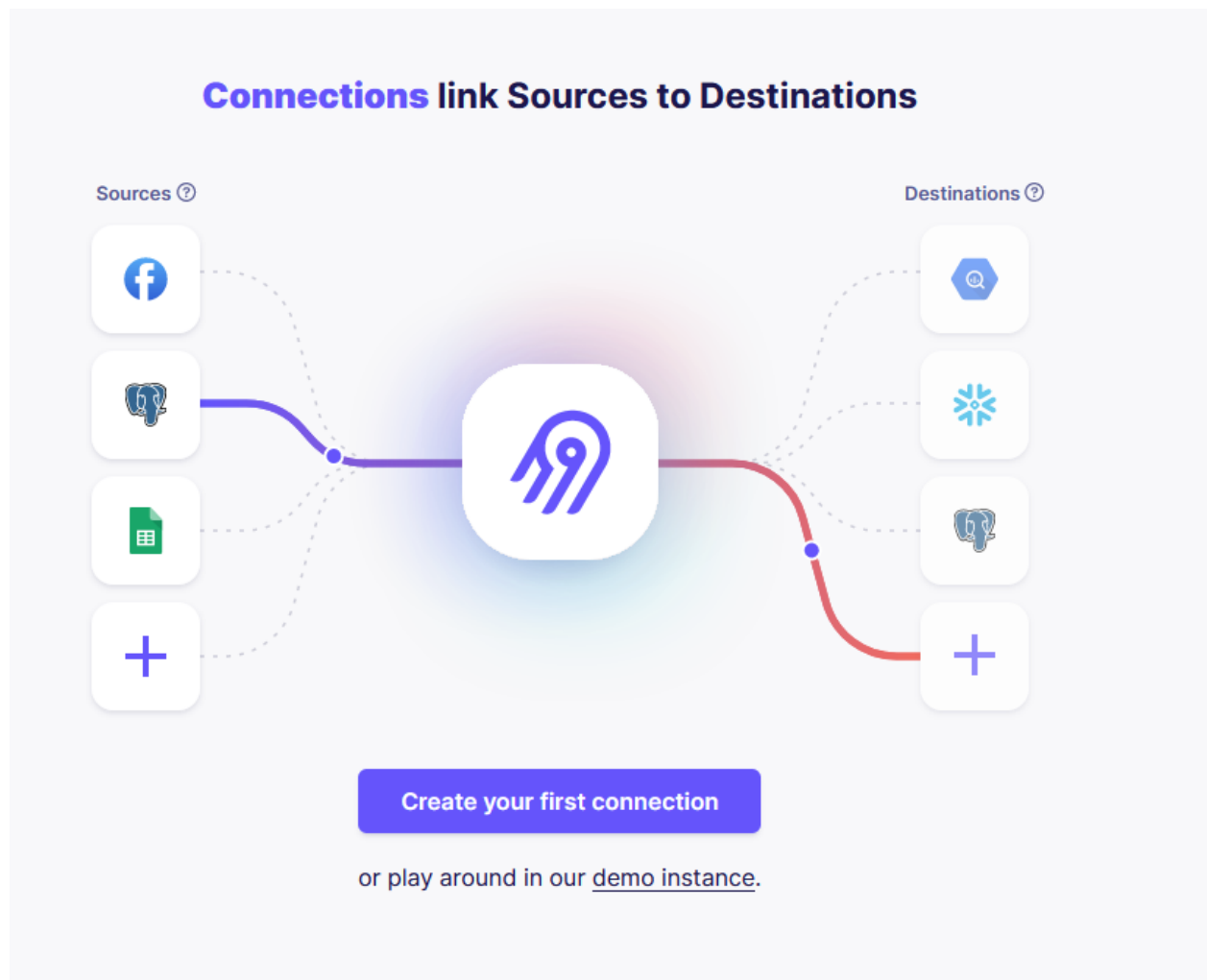
```
Building 0.63 (0/0)
Running 1/1
Network airbyte_airbyte_internal Created
Network airbyte_airbyte_public Creating
Running 2/5me "airbyte_workspace" already exists but was not created for project "airbyte-project". Use `external: true` to use an existing volume
Network airbyte_airbyte_internal Created
Network airbyte_airbyte_public Created
Container airbyte-temporal Creating
Container airbyte-db Creating
Container init Creating
```

Una vez desplegado, nos dirigimos al navegador con la siguiente url:

<http://localhost:8000/>

Que es donde se encuentra nuestro servicio.

Allí, nos vamos a encontrar con la interfaz gráfica donde podemos crear nuestra fuente de datos, el destino, y la conexión, que es el puente entre la fuente y el destino.



Casos de uso

Airbyte se utiliza principalmente para integrar y consolidar datos provenientes de diversas fuentes en un único lugar. Algunos casos de uso comunes de Airbyte son:

- **Integración de datos:** Airbyte facilita la extracción y carga de datos desde diferentes fuentes, como APIs, bases de datos y aplicaciones. Esto es útil para empresas que manejan múltiples fuentes de datos dispersas y desean centralizarse en un único repositorio.
- **Análisis de datos:** Al tener los datos consolidados en un solo lugar, permite realizar análisis y exploración de datos de manera más eficiente. Los datos pueden ser transformados y modelados para adaptarse a las necesidades específicas de análisis de la empresa.
- **Sincronización de datos:** Es posible mantener actualizados los datos en tiempo real o programar sincronizaciones periódicas. Esto garantiza que los datos estén siempre actualizados y disponibles para su uso.
- **Data Warehousing:** Es útil para cargar datos en bases de datos de big data (conocidas como data warehouses) o en plataformas de big data, como Amazon Redshift, Google BigQuery o Snowflake. Esto permite tener una visión unificada de los datos y facilita su posterior análisis y consulta.

Contribución

Al ser una herramienta open source, permite contribuciones en su repositorio de GitHub, allí se pueden agregar conectores, contribuir con errores no solucionados y abrir errores que uno encuentra.

<https://github.com/airbytehq/airbyte>

Laboratorio

Para el laboratorio, se implementó la herramienta Airbyte utilizando su repositorio como punto de partida.

En otro escenario, se incluyó una base de datos para simular un entorno de data warehouse y probar diferentes conexiones dentro del mismo entorno docker compose. Aunque PostgreSQL no es la herramienta más óptima para este propósito, se utilizó como solución de simulación. Además, se incorporó Adminer, un gestor de contenido de base de datos, para manipular y verificar los datos del servidor PostgreSQL.

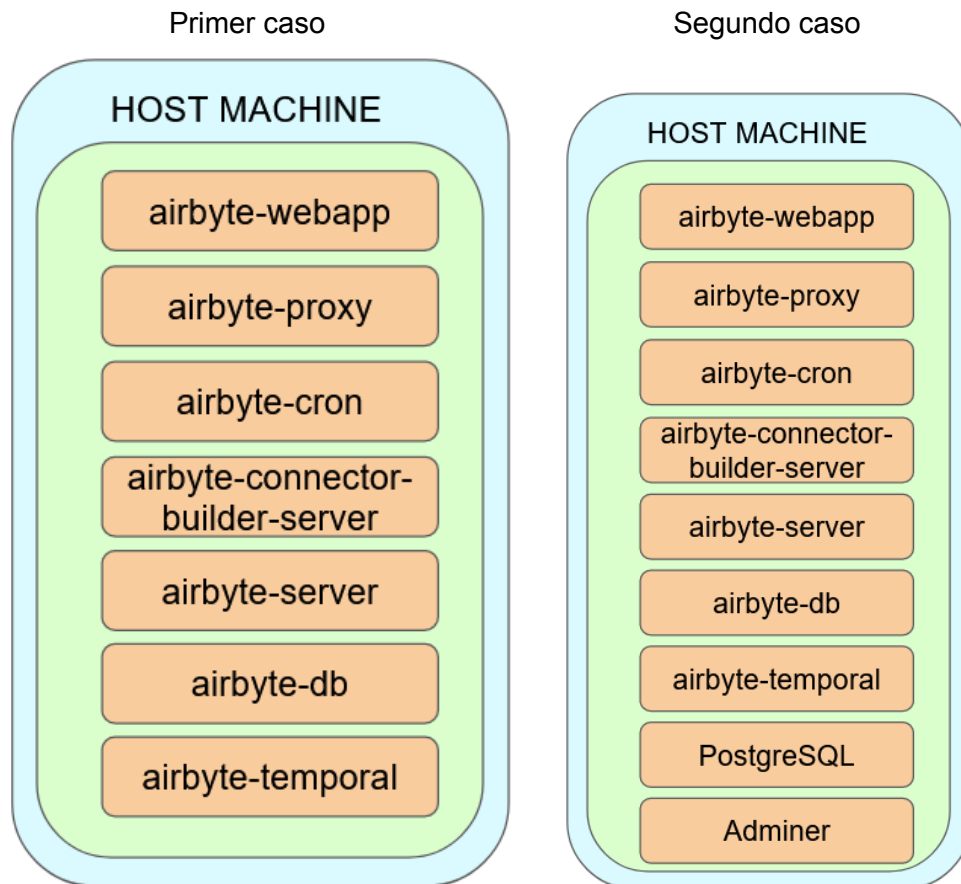
Se pueden observar los diferentes casos en este repositorio:

<https://github.com/Nahuel-P/unq-labo-so>

El primer caso se encuentra en el archivo `docker-compose.yaml`

El segundo caso con la base de datos se encuentra en `docker-compose-modified.yaml`

Ambas arquitecturas quedan de la siguiente manera:



Para ejecutar diferentes docker compose dependiendo del caso, se ejecuta los siguientes comandos:

```
cd 1st_scenario # or cd 2nd_scenario
docker-compose up
```

Referencias

<https://docs.docker.com/engine/install/>

<https://docs.docker.com/compose/install/>

<https://github.com/airbytehq/airbyte>

<https://docs.airbyte.com/>