



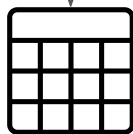
Amazon
EC2



snowpipe



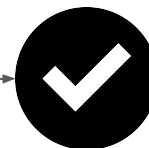
Amazon
S3



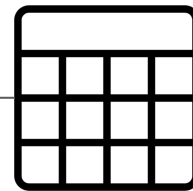
Staging Table



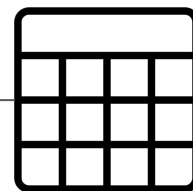
Snowflake Stream



Snowflake Task



Target Table 2



Target Table 1

Apache Nifi

What is Apache NiFi?

- A data integration tool.
- Created by the National Security Agency (NSA) and later contributed to the Apache Software Foundation.
- Provides a web-based interface for designing, controlling, and monitoring data flows.

Key Concepts:

- FlowFile: The basic unit of data in NiFi. It has content (actual data) and attributes (metadata).
- Processor: A node that performs an operation on a FlowFile, e.g., fetching data, transforming data, routing data based on conditions.
- Connection: Links processors together, allowing FlowFiles to move from one processor to another.
- Flow Controller: The brains of the operation, ensuring threads are allocated to tasks.
- Process Group: A sub-flow or a collection of processors and connections. Helps in organizing complex flows.

Apache Nifi

Why Use NiFi?

- Visual Design: Drag-and-drop interface for designing data flows.
- Scalability: Can handle large amounts of data and is horizontally scalable.
- Flexibility: Supports various data sources and destinations (e.g., databases, HTTP, FTP, Kafka, Hadoop).
- Data Provenance: Tracks data from its source to destination, allowing for auditing and troubleshooting.
- Extensibility: Custom processors can be developed if needed.

0 0 / 0 bytes 0 0 0 0 0 359 2 52 0 0 0 0 0 0 16:37:44 CEST

Search

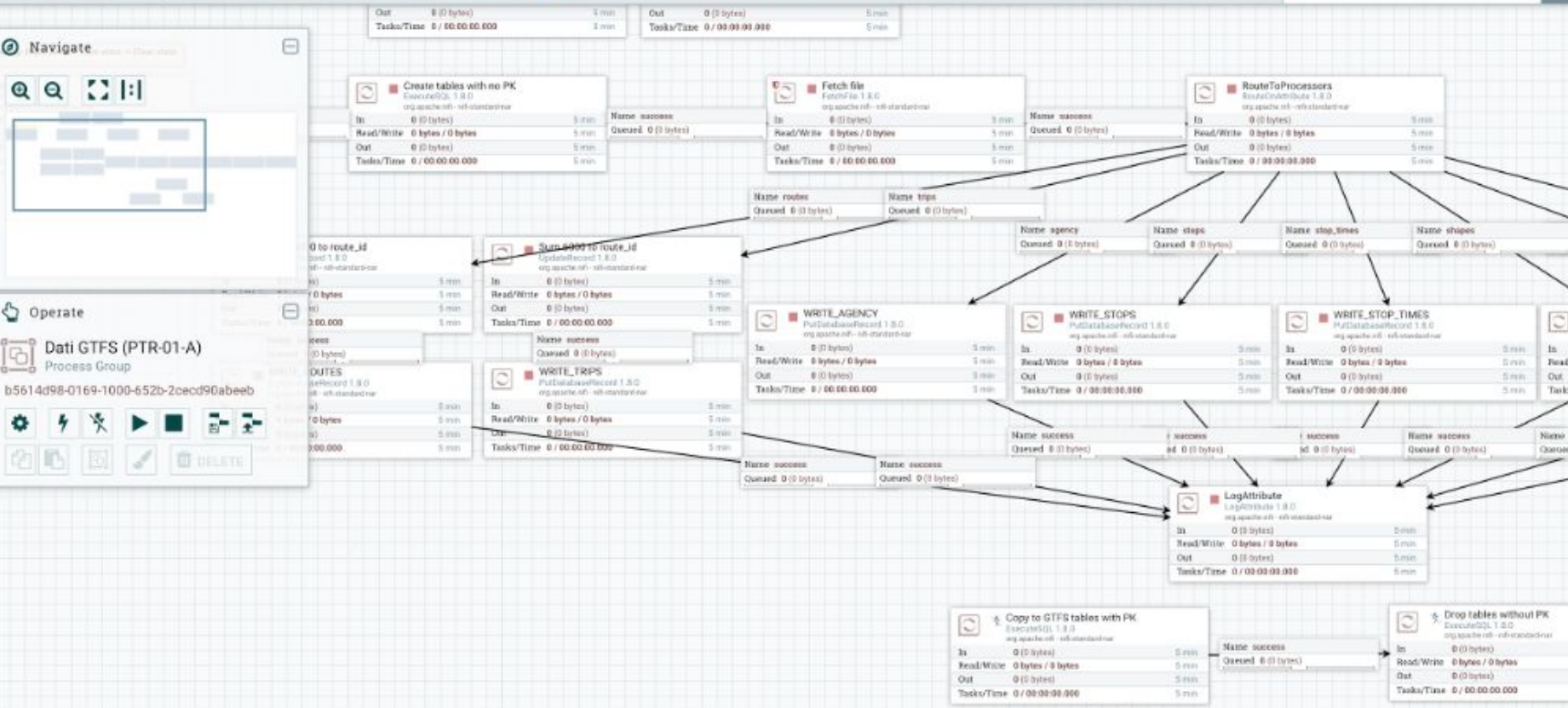
Navigate



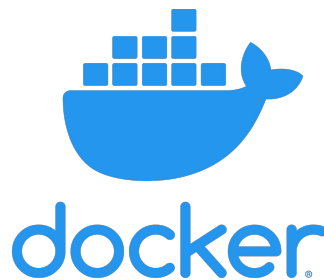
Operate

Dati GTFS (PTR-01-A)
Process Group

b5614d98-0169-1000-652b-2cecd90abeeb



Docker



Docker Images:

- Simple Explanation: It's like a blueprint or a recipe.
- Details: An image is a lightweight, stand-alone package that contains everything needed to run a piece of software, including the code, runtime, libraries, and other dependencies.

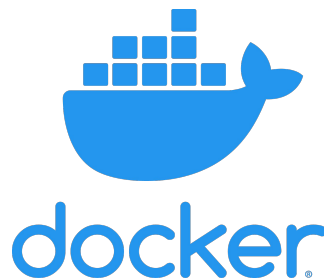
Docker Containers:

- Simple Explanation: It's the actual box we talked about earlier. It's created from an image.
- Details: A container is a runnable instance of an image. You can think of it as the live version of the Docker image.

Docker Hub:

- Simple Explanation: It's like a library or store where people share their blueprints (images) so others can use or modify them.
- Details: Docker Hub is a cloud-based registry where you can find and share container images with your team and the Docker community.

Docker



Dockerfile:

- Simple Explanation: It's a set of instructions to create an image.
- Details: A Dockerfile is a script with commands to assemble a Docker image. For example, it might specify a base image, add files, and set some configurations.

Docker Compose:

- Simple Explanation: If you have multiple toys (applications) that play together, you use Docker Compose to organize and run them together.
- Details: Docker Compose is a tool for defining and running multi-container Docker applications. You define everything in a docker-compose.yml file and then use the docker-compose command to start the entire stack.