***Steps for Deploying Talend job into Snowflake using Airflow.***

1. **Setting up Snowflake Account:**

* Created a new snowflake account:
* Snowflake:

Account: [https://yab67224.us-east-1.snow\*\*\*\*\*\*\*/](https://yab67224.us-east-1.snow*******/)

UserName: Khalidairflow

Password: Prometric@2024

warehouse= COMPUTE\_WH

database= AIRFLOW\_SNF

schema= AIRFLOW\_SCHEMA

Region: US East (N. Virginia)

Role: ACCOUNTADMIN

**Created Structure, Stage, and file format in Snowflake:**

create or replace TABLE AIRFLOW\_SNF.AIRFLOW\_SCHEMA.AIRFLOW\_TABLE (

ID NUMBER(38,0),

NAME VARCHAR(16777216),

COMPANY VARCHAR(16777216)

);

create or replace file format my\_csv\_format

type = csv field\_delimiter = ',' skip\_header = 1

field\_optionally\_enclosed\_by = '"'

null\_if = ('NULL', 'null')

empty\_field\_as\_null = true;

CREATE OR REPLACE STAGE AIRFLOW\_SNF.AIRFLOW\_SCHEMA.AIRFLOW\_TABLE\_stage

URL = 's3://s3stag/landing\_directory/'

CREDENTIALS = (

AWS\_KEY\_ID = 'AKIATNIWVCQAYAQTXCPS',

AWS\_SECRET\_KEY = 'INY4AN472h\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*’

)

FILE\_FORMAT = my\_csv\_format;

show stages;

list @AIRFLOW\_SNF.AIRFLOW\_SCHEMA.AIRFLOW\_TABLE\_stage;

select \* from AIRFLOW\_SNF.AIRFLOW\_SCHEMA.AIRFLOW\_TABLE;

A screenshot of a computer

Description automatically generated

1. **Setting up Talend Job**

**Develop Talend Job:**

* Created a new talend job to deploy into snowflake as shown below.

A screenshot of a computer

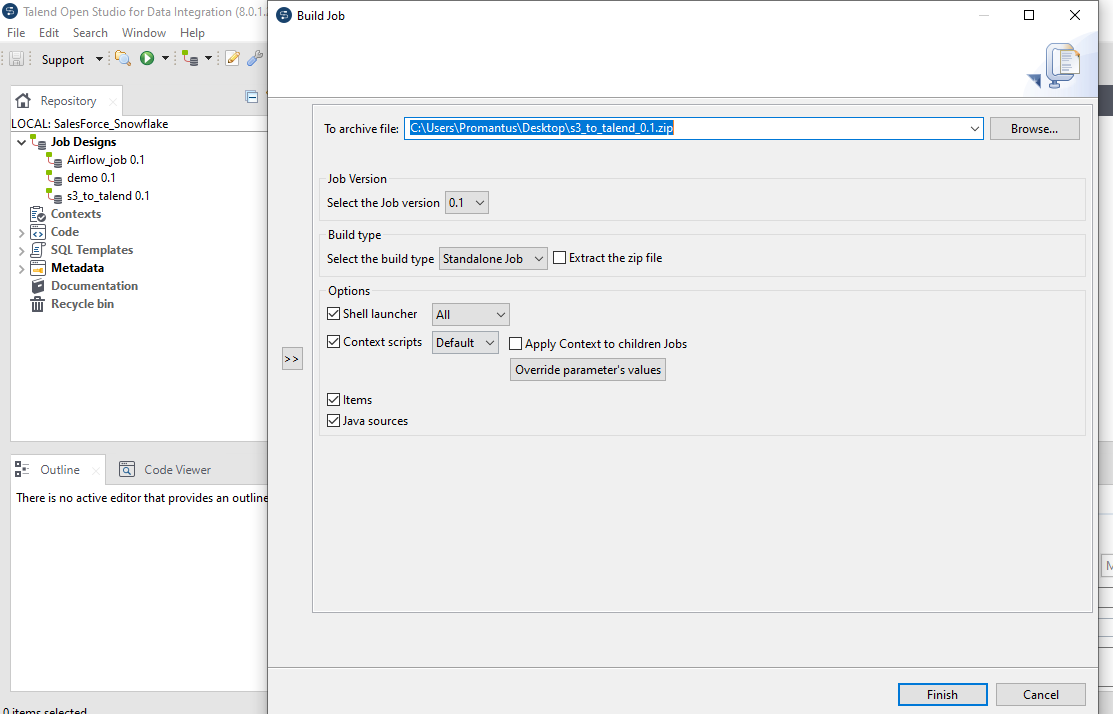
Description automatically generated

* In metadata set connection to snowflake using credentials.
* tDBRow\_1: Type below code to load data into snowflake.

"copy into AIRFLOW\_SNF.AIRFLOW\_SCHEMA.AIRFLOW\_TABLE

From @AIRFLOW\_SNF.AIRFLOW\_SCHEMA.AIRFLOW\_TABLE\_stage FILE\_FORMAT = (FORmaT\_NAME =my\_csv\_format)"

* Build standalone talend job as show in below image.



* placed the job zip file in s3 bucket **talend-to-snf/codebase.**

1. **Setting up Apache Airflow on AWS EC2:**

* Launch an EC2 instance on AWS with proper role and required packages to operate Talend, Python operator, snowflake, and airflow as follows.
* Install openjdk-8-jdk since Talend requires.
* Install unzip to extract build job-name.zip from Talend.
* Install virtual machine to configure Airflow to use an external metadata database (e.g., PostgreSQL, MySQL)
* Install Apache-airflow [Postgres]==2.6.0
* Set user Id and password for login to airflow using public Ip address of ec2.
* Install python3.
* Install Apache-airflow-providers-snowflake. ...etc.
* The details are attached.
* Create S3bucket give full accesses for the role you created while launching ec2.
* Create log folder in S3 bucket: where you capture all logs of the packages.
* Create codebase folder in S3 bucket: where you kept your DAG and job-name.zip.
* Create [landing directory](https://s3.console.aws.amazon.com/s3/buckets/talend-to-snf?region=ap-south-1&bucketType=general&prefix=landing_directory/&showversions=false) folder in S3 bucket: where you need to put csv and success file.

Creating S3 Bucket

A screenshot of a computer

Description automatically generated

Codebase Landing directory log

A screenshot of a computer

Description automatically generated A screenshot of a web page

Description automatically generated A screenshot of a computer

Description automatically generated

**Launch and Connect Instances**

A screenshot of a computer

Description automatically generated

**Connect Terminal and execute the code to connect to airflow.**

A screenshot of a computer

Description automatically generatedA screenshot of a computer program

Description automatically generated

**Giving Public IP address of Ec2 follow by 8080 port of Airflow using Airflow user id and password which we define under packages Login to Airflow**

A screenshot of a computer

Description automatically generated

1. **Configuring Airflow Connections:**

* Launch Airflow using Ip address of ec2.
* Configure Snowflake connection in Airflow's Connection settings with the Snowflake credentials obtained earlier.
* Configure other necessary connections for AWS (e.g., S3 staging)

**In Airflow the Dag you loaded will be executed**

A screenshot of a computer

Description automatically generated

**Connect to AWS Connect to Snowflake**

A screenshot of a computer

Description automatically generated A screenshot of a computer

Description automatically generated

1. **Creating an Airflow DAG:**

* Define an Airflow Directed Acyclic Graph (DAG) that includes tasks for executing your Talend job.
* Use the Bash Operator or Python Operator to execute your Talend job as a task within the DAG.
* Define any necessary dependencies between tasks.

1. **Start Airflow Schedular:** Ensure that the Airflow scheduler is running so that it can schedule and execute your DAGs according to the specified schedule.
2. **Monitor Execution:** Monitor the execution of your DAG in the Airflow UI. Check logs for any errors or issues that may arise during execution.
3. **Maintenance and Update:** Regularly maintain and update your Airflow environment, DAG scripts, and Talend jobs to ensure they continue to run smoothly and efficiently.