



Data Engineering Team - Take-home Test

From Recruitment Cake <recruitment@cake.vn>

Date Tue 12/30/2025 10:14 AM

To nxthuong@hotmail.com <nxthuong@hotmail.com>

Cc Trần Thị Thanh Trúc (TABP Manager) <truc.tran2@cake.vn>

Dear Nguyễn Xuân Thường,

Thank you for your interests in Cake's opportunities.

At the first session of Cake's hiring process, we would like to invite you to complete the Take-home Test from our Data Engineering Team. Please spend time doing the test and reply your submission within 5 days. If you have any further questions or require any changes, please do not hesitate to contact us for discussion.

Problem Statement

The task involves two SFTP destinations, referred to as <source> and <target>.

Your objective is to develop an Apache Airflow DAG that facilitates the transfer of files from the SFTP server at <source> to the SFTP server at <target> and ensures the preservation of the original directory structure.

The synchronization process should be unidirectional; hence, any modification made on <target> must not impact the <source>.

Deleted files on SFTP server at <source> must remain intact on <target> server.

Examples:

- On March 1st, 2024, when a file named sftp://<source>/a/b/c/file_1.txt is detected on the source server, it should be replicated to sftp://<target>/a/b/c/file_1.txt on the destination server.
- On March 2nd, 2024, a file named sftp://<source>/a/b/c/file_2.txt appears on the source server and subsequently should be transferred to sftp://<target>/a/b/c/file_2.txt on the destination server.

- On March 3rd, 2024, a file named sftp://<source>/a/b/c/file_3.txt appears on the source server and should then be transferred to sftp://<target>/a/b/c/file_3.txt on the destination server.

Expected Outcome

Candidates are expected to provide a **GitHub** public repository as following:

- Use separated commits that reflect your incremental development and refactoring. Pure atomic commits are not expected, and don't squash them.
- A docker-compose.yml file for deploying latest Airflow version, with each service (Scheduler, Worker, Web Server) running in a separate container. Utilizing SQLite as the backend is permissible. The use of Celery executor is advised.
- A README.md file that includes:
 - Detailed instructions for setting up and running the project.
 - Any additional initial setup requirements.
 - An explanation of assumptions made, decisions taken, and any trade-offs considered.
 - One or more DAG file(s).
 - Any additional plugins required for the project.

What We Don't Expect

A comprehensive, production-ready ETL/ELT solution that is agnostic to data sources and destinations

What We Assess

- The functionality of a runnable Airflow DAG that accurately achieves the specified result.
- The candidate's adherence to a consistent and clean coding style.
- The level of abstraction in your API(s). Given that business requirements are subject to change, evaluate how adaptable your solution is if the data source transitions from SFTP to Object Storage.
- The extensibility of your API(s). Assess whether it is feasible to incorporate additional transformations before loading files into the target system without significant effort.
- Your strategy for handling anomalies. For instance, if file sizes increase dramatically from kilobytes to gigabytes, how does your solution accommodate this change in scale?

With warm regards!

Thanks & Best Regards,



Talent Acquisition Business Partner Team

Cake - Human Resources Division

Email: recruitment@cake.vn

Website: <https://cake.vn/tuyen-dung>