

Chapitre 4 :

Airflow

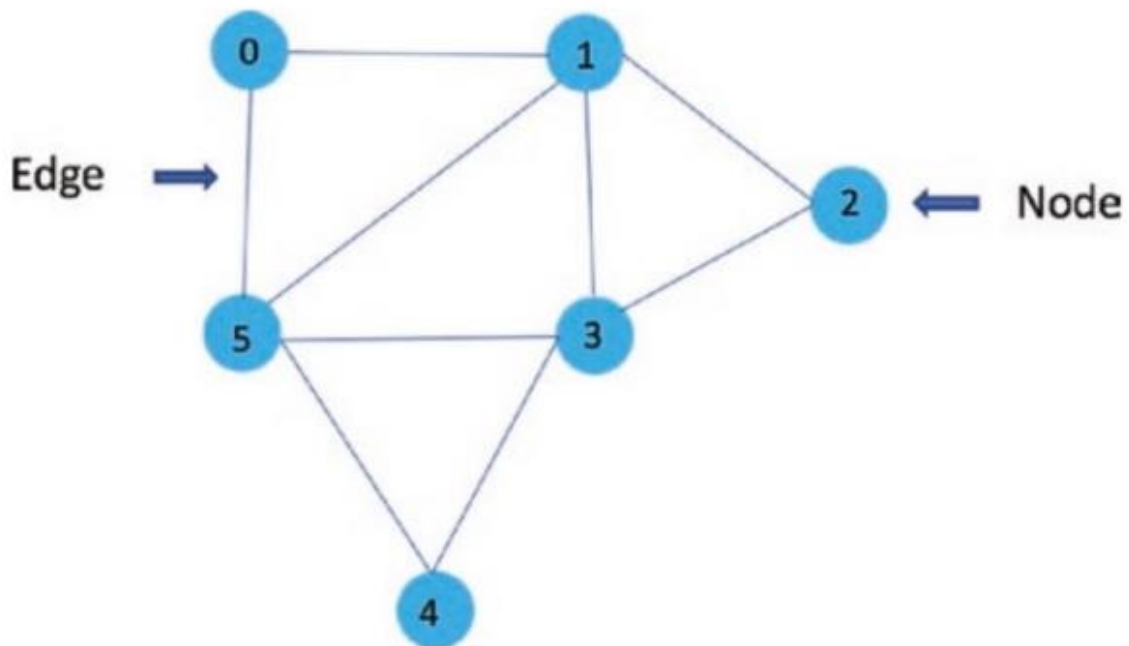
A. Notes de cours

Un flux de travail doit être conçu et exécuté, afin d'obtenir les résultats souhaités. Comme sur la figure ci-dessous



Une structure de données graphique typique se compose de deux entités :

1. Bords
2. Nœuds/sommets



Les arêtes sont essentiellement les connexions entre les nœuds/sommets, et les nœuds sont l'endroit où résident les données réelles.

Les deux catégories principales des réseaux basés sur des graphes

1. Graphes non orientés : Dans ce type de structure de graphe, les arêtes ou les connexions n'ont aucune direction. La relation existera aux deux extrémités

2. Graphes orientés : est soit cyclique(ont un ou plusieurs cycles), soit acyclique. La direction de l'arête joue un rôle important, car toutes les arêtes d'un graphe ne vont que dans un sens

- ✓ Un cycle est un chemin qui commence et se termine au même nœud.
- ✓ Un DAG est un graphe orienté, les informations ne peuvent circuler que dans une seule direction, et c'est vers l'avant. Donc il n'y a pas de chemin inverse vers le nœud de départ.
- ✓ Toutes les tâches des jobs dans Airflow doivent être définies dans un DAG. Ainsi, l'ordre d'exécution est défini sous forme de DAG.
- ✓ Toutes les configurations liées au DAG sont définies dans un fichier de définition DAG, qui est une extension Python. Il contient toutes les dépendances et les paramètres de configuration, tels que l'e-mail à envoyer en cas d'échec, l'heure de début, l'heure de fin et le nombre de tentatives.

✓ Les opérateurs les plus courants sont :

1. Opérateur Python (script Python)
2. Opérateur bash (script shell)
3. Opérateur SQL
4. Opérateur Docker
5. Opérateur Cloud (S3, Azure, Google)