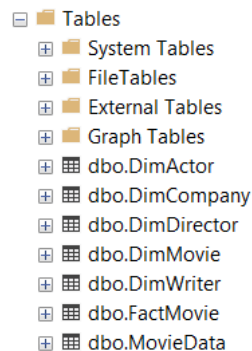


Création et remplissage des tables l'entrepôt de données

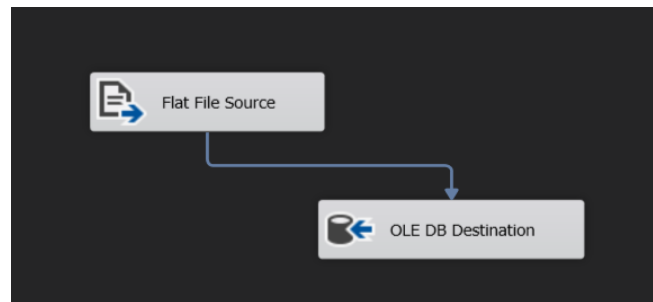
Création de l'entrepôt de données

L'entrepôt de données est créé localement sur la machine (localhost) dans Microsoft SQL Server et est géré par le biais de Microsoft SQL Server Management Studio 18 (SSMS). Il faut tout d'abord créer les tables de dimensions et la table de fait qui vont contenir les données à analyser, et une table MovieData, qui va servir de table intermédiaire afin de remplir la table de fait. La création des tables et les commandes de remplissage sont contenues dans le fichier requetesSQL.sql. Les tables peuvent être visualisées dans le dossier Tables de la base de données de SSMS :



Remplissage des tables

La première table à remplir est la table MovieData qui contient les données brutes et les clés de chaque table. Le remplissage est fait par le biais de l'outil ETL Microsoft SQL Server Integration Service (SSIS) dans Microsoft Visual Studio 19. Les données sont lues directement depuis le fichier csv et importées dans la Base de données.



Une fois les données importées, pour chaque table de dimensions, nous allons insérer les tuples contenues dans MovieData en faisant attention d'utiliser 'DISTINCT' dans la requête afin d'éviter les répétitions dans les tables. Les commandes d'insertion sont contenues dans le fichier requetesSQL.sql. Par exemple, la table de dimensions des acteurs est remplie avec les noms des acteurs et un identifiant leur sont automatiquement associés. Le remplissage est répété pour chaque table de dimension.

La table MovieData doit ensuite être mise à jour avec les identifiants des tables de dimensions, les requêtes de mise à jour sont contenues dans le fichier SQL. Après avoir rempli cette table, la table de fait peut ensuite être remplie en sélectionnant les tuples de la table MovieData.