

MU5IN852
Bases de Données
Large Echelle

TME 1 :
analyse de
la base IMDB

septembre 2021

hubert.naacke@lip6.fr

TME sur la base de films IMDB

- Explorer des fichiers de données
 - Fichier vers table
- (re)structurer les données
 - Pour une analyse :
 - Lister les attributs analysés
 - Structure du résultat des analyses
- Définir les transformations de manière déclarative
 - en SQL

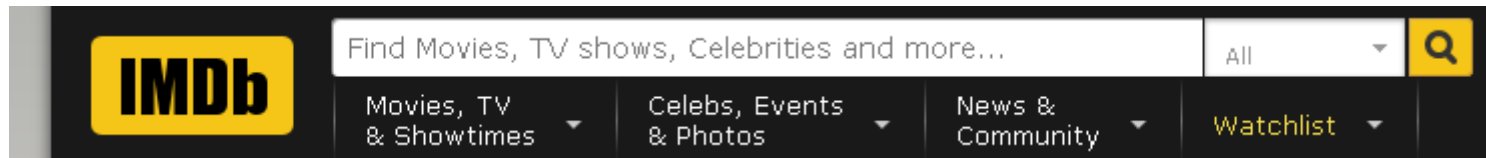


Schéma de la base imdb

- Les personnes : Name, Person_Info,
- Les films : Title, Movie_Info
- Les sociétés de production : Company_name
- Les associations :
 - film/personne : Cast_Info
 - film/société : Movie_companies

Schéma d'une **personne**

- Une personne : Name, Person_Info
- Casting : une personne a un ou plusieurs rôles dans un film

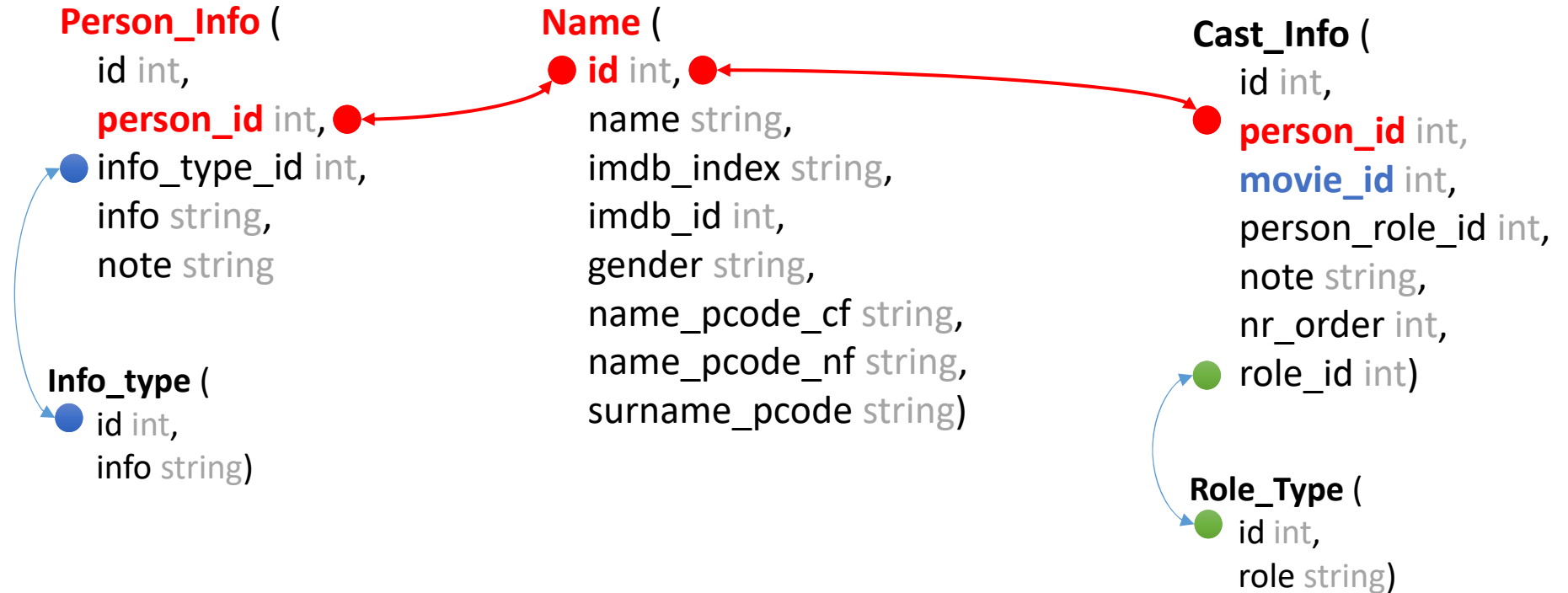


Schéma d'un film

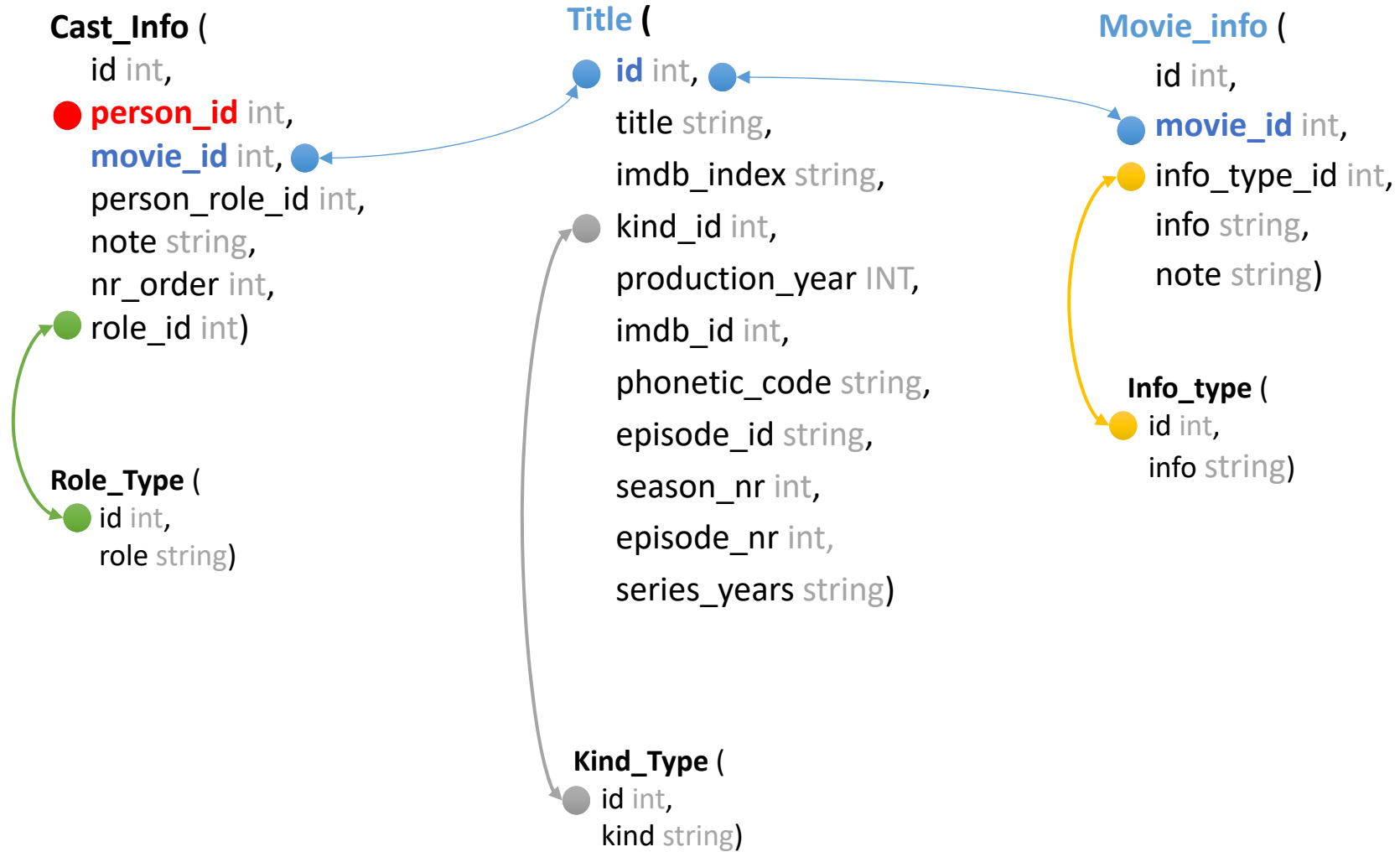
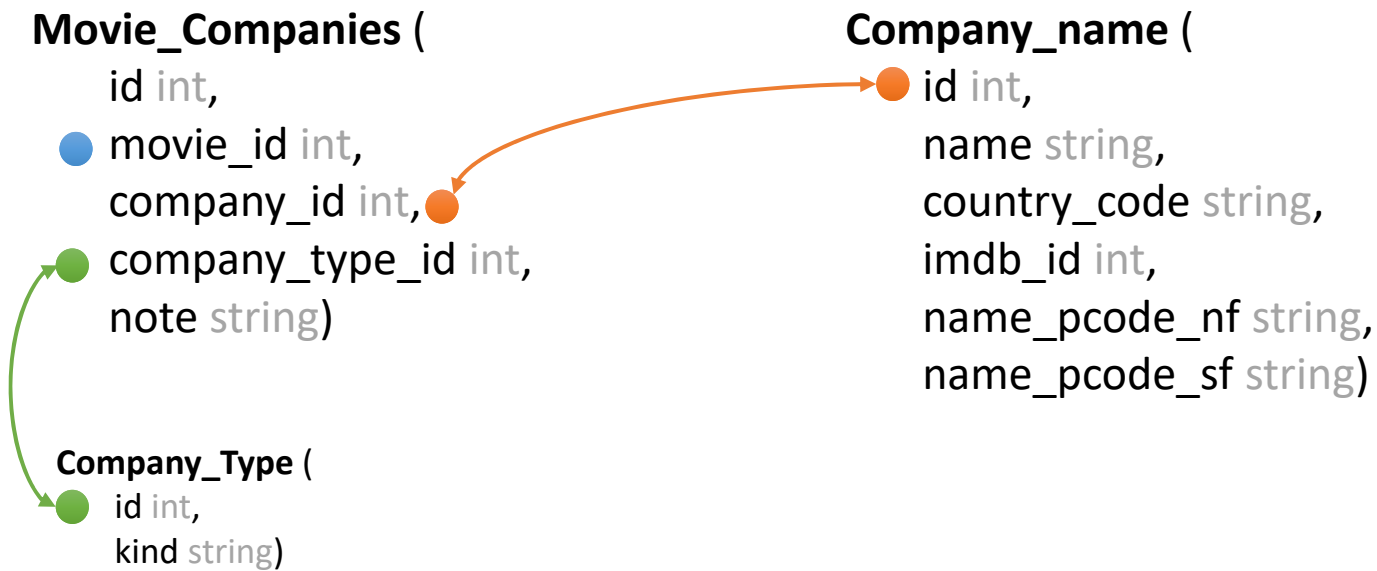


Schéma d'une société de production



Datasets

- Dossier partagé : PUBLIC_DATASET
<https://nuage.lip6.fr/s/H3bpyRGgnCq2NR4>
- Fichiers IMDB dans le dossier
 - imbd/vldb2015/csvfiles_sample001
- Contient aussi d'autres datasets à disposition