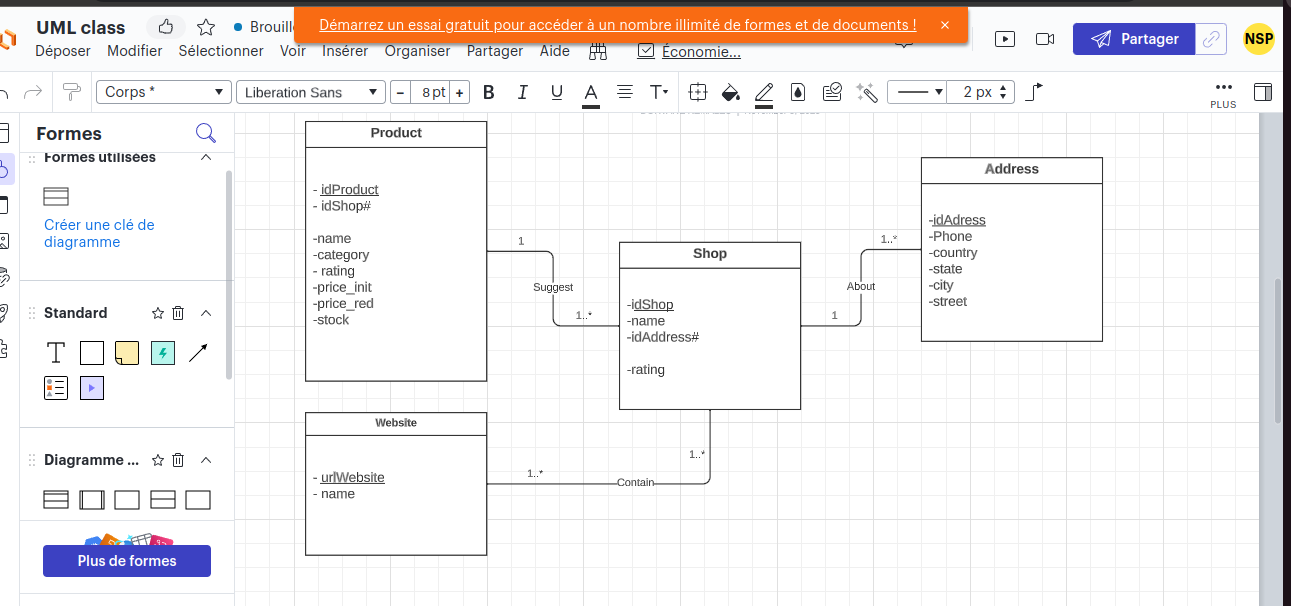


PROJECT

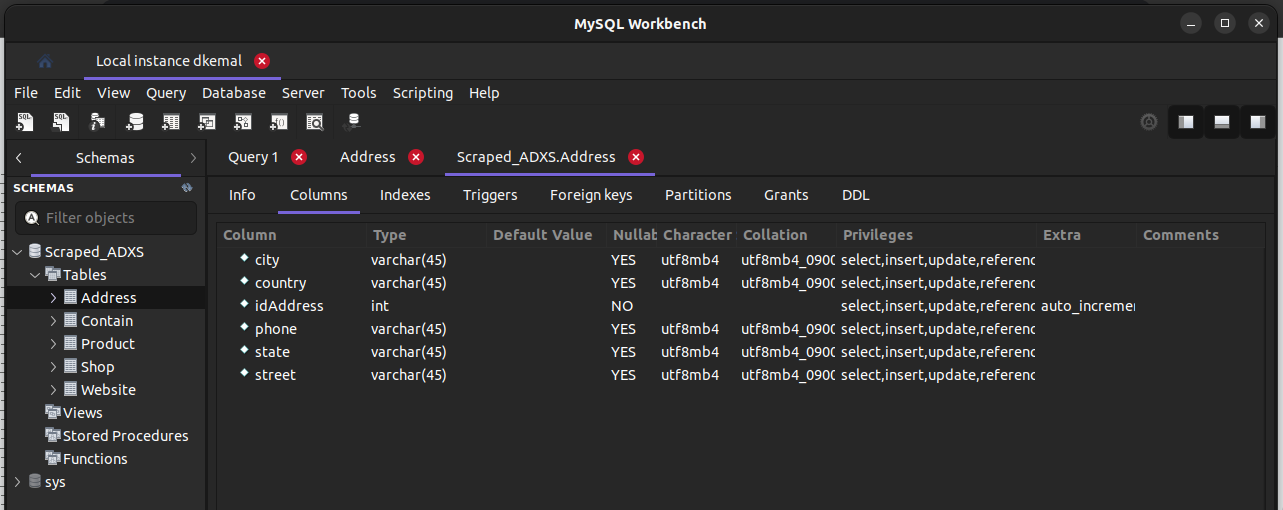
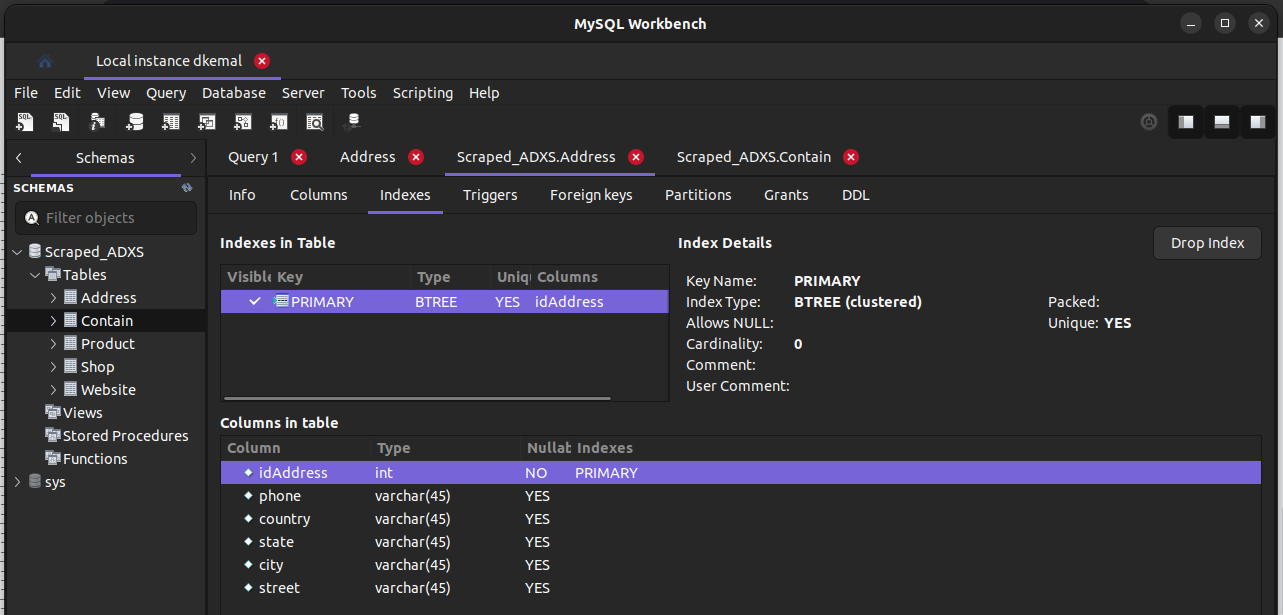
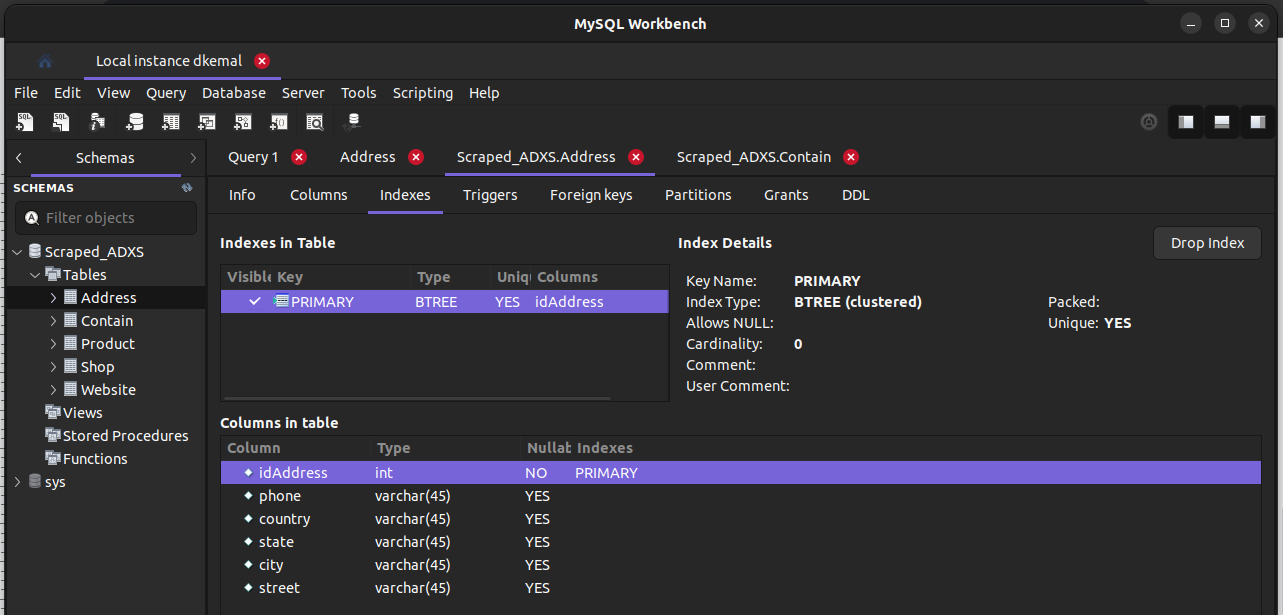
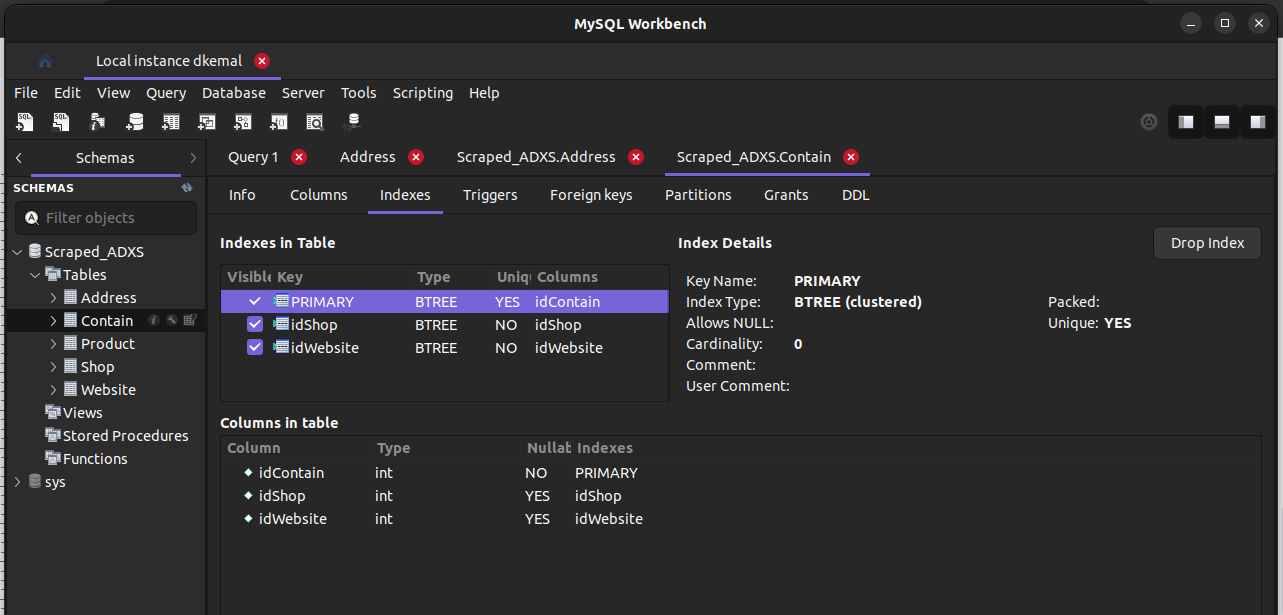
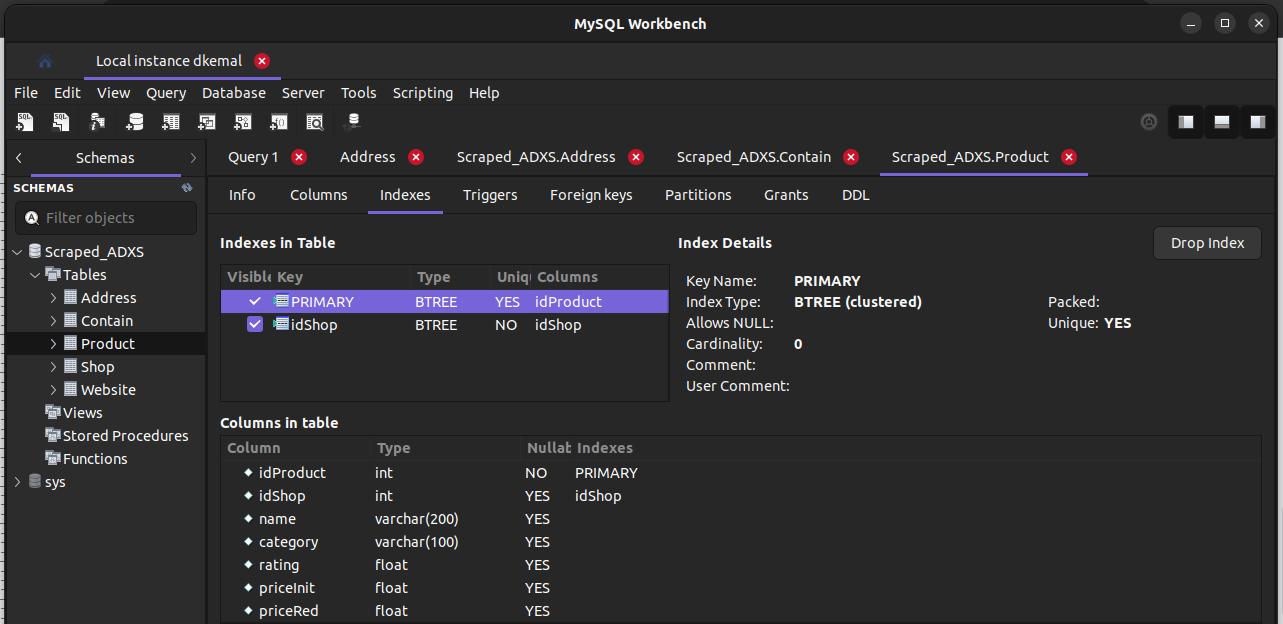
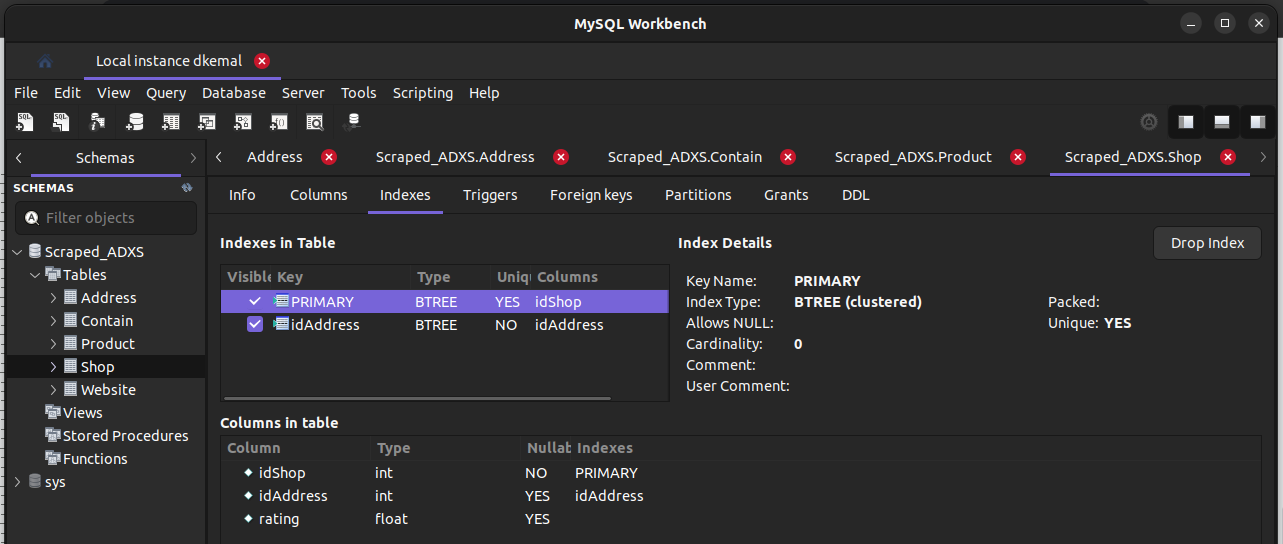
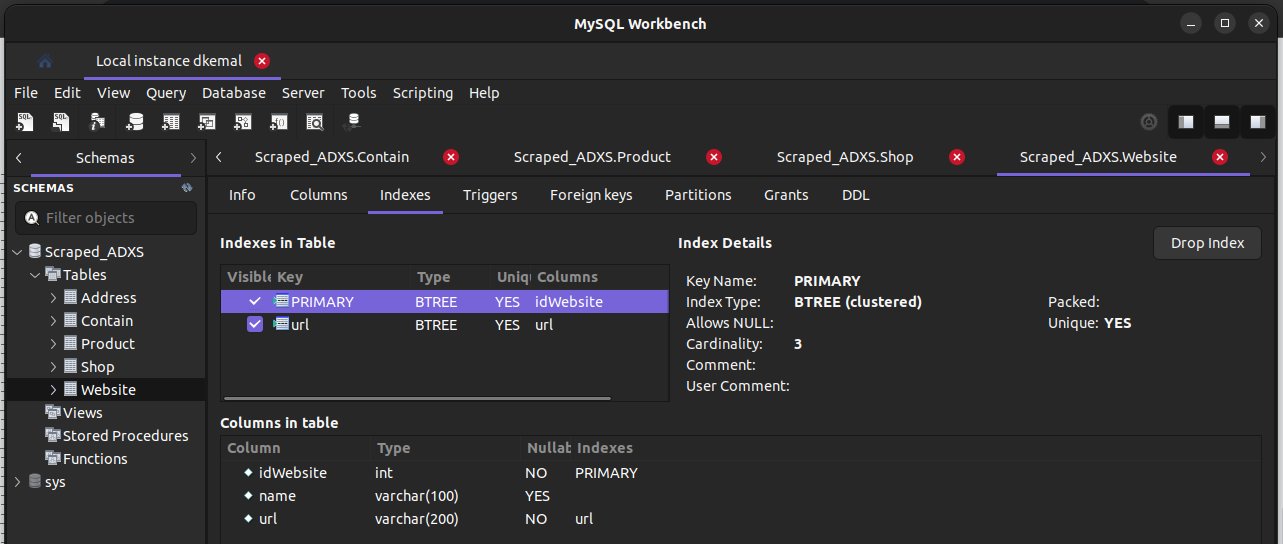
**E-COMMERCE DATA SCRAPING & ANALYTICS**

**MODELISATION DES DONNEES**

**Modélisation conceptuelle**



**Modélisation logique en considérant un SGBD SQL : MySql**

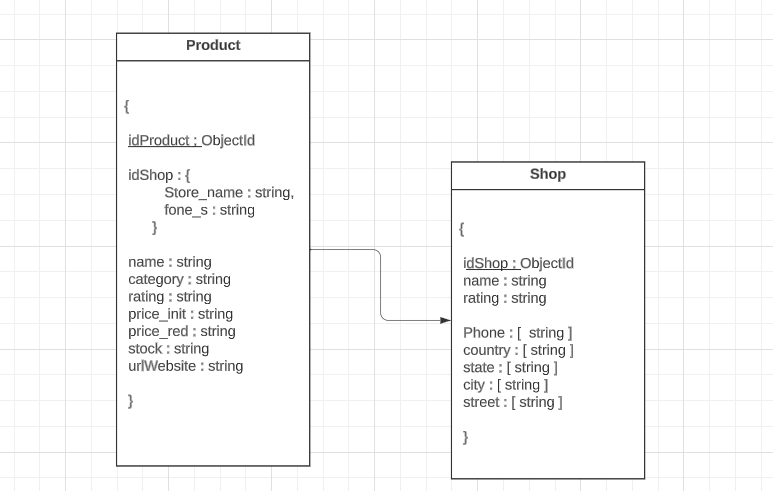


**Avantages et limites de la bd SQL**

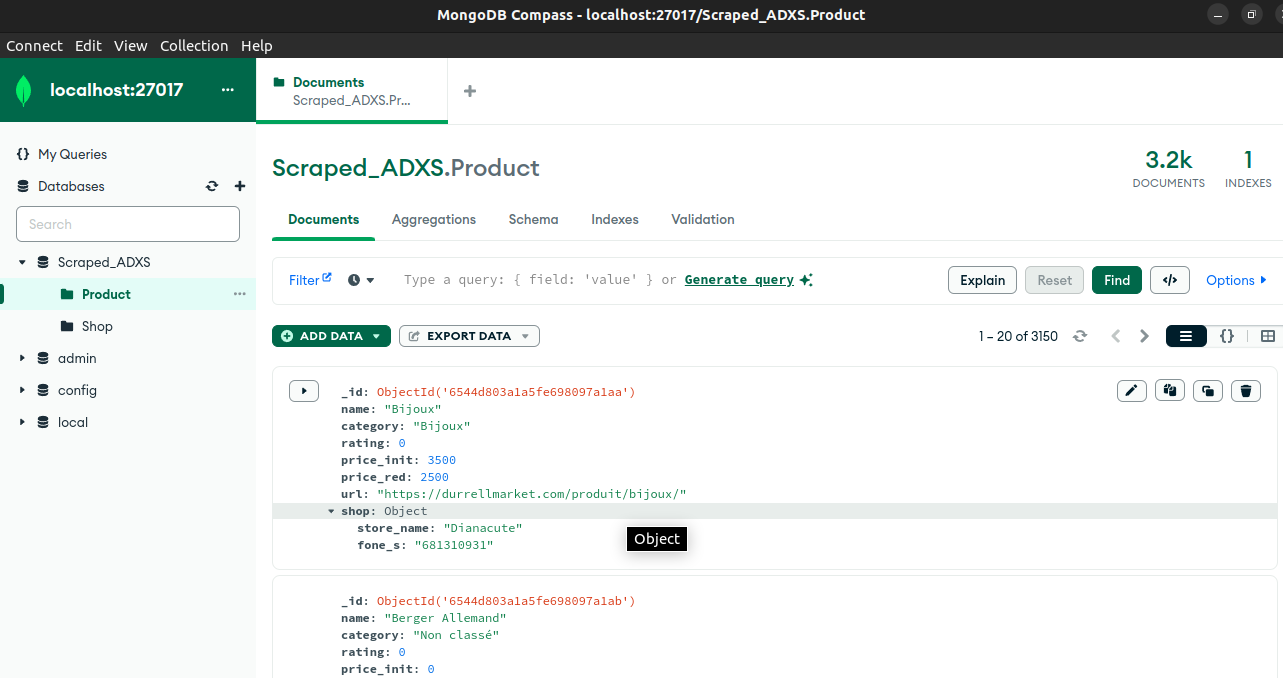
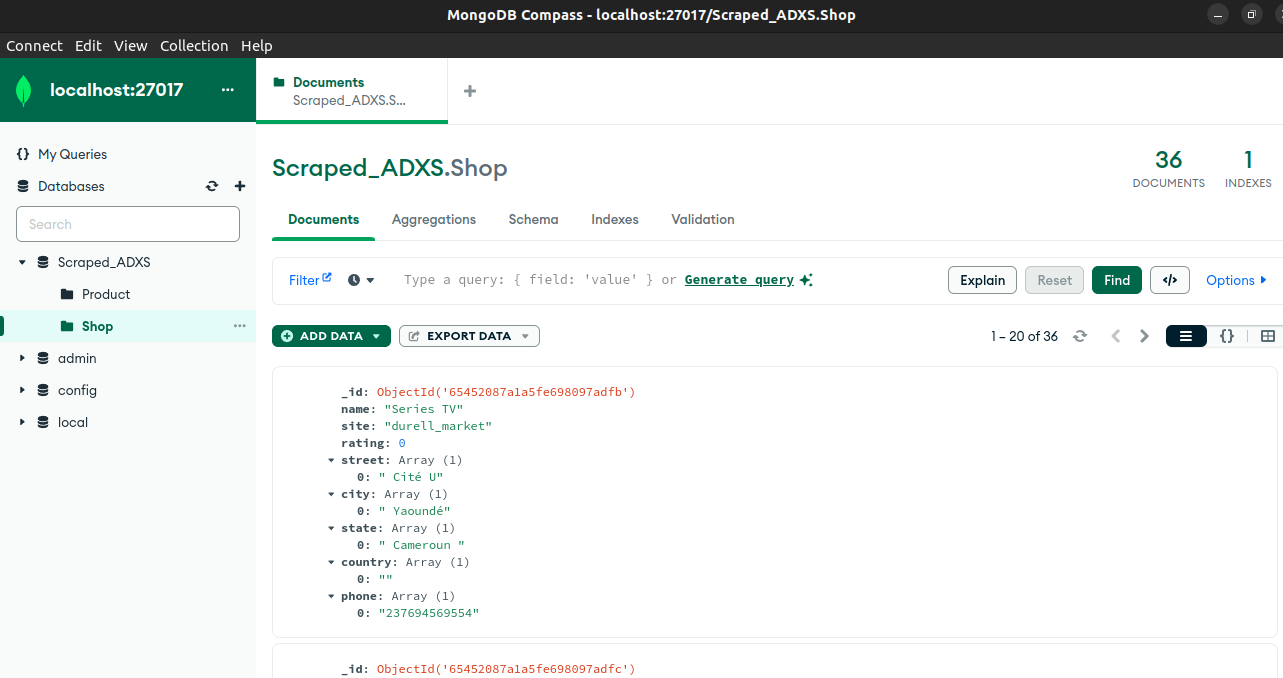
Cette base de données est assez simple à concevoir et monter, ACID (Atomicité, Cohérence, Isolation, Durabilité). Mais elle est très rigide, peu évolutive et inadaptée pour les Big Data. Une simple opération de mise à jour/ chargement pourrait être très coûteuse.

**Modélisation logique en considérant un SGBD NoSQL Document : MongoDB**

**Visualisation des Collections**



**Base de données**



**Avantages et limites de cette Base de donnée NoSQL**

La logique de ce modèle est très proche de nos données scrappées. La bd est plus flexible, et performante pour des requêtes simples en temps réel. Elle supporte une scalabilité horizontale contrairement à la précédente, et est idéale pour nos Big Data.

**DICTIONNAIRE DE DONNÉES**

| **Collection / Table** | **NomChamp** | **Type Champ** | **Description** |
| --- | --- | --- | --- |
| **Poduct** | **id** | ObjectId | Identifiant du document “Produit”, généré et traité par le SGBD |
| **name** | String | Nom du produit |
| **category** | String | Catégorie à laquelle appartient le produit |
| **rating** | Float | Moyenne des notation des clients qui apparaît sur le site. elle est notée sur 5 |
| **price\_init** | Float | Prix du produit (barré), sans la réduction qui apparaît sur le site |
| **price\_red** | Float | Prix final du produit ( non barré) qui apparaît sur le site |
| **url** | string | Url qui identifie le site sur lequel le produit à été récupéré |
| **shop** | Object | Objet qui contient les premières informations apparentes (store\_name, fone\_s) de la boutique qui propose ce produit |
| **store\_name** | string | Nom de la boutique qui propose le produit |
| **fone\_s** | string | Numéro de téléphone de la boutique qui propose le produit |
| **Shop** | **id** | ObjectId | Identifiant du document “Produit”, généré et traité par le SGBD |
| **name** | string | Nom de la boutique |
| **site** | string | Nom du site sur lequel elle a été récupérée |
| **rating** | string | Moyenne de notation de la boutique sur le site ou elle a été récupérée |
| **city** | [ string ] | Liste des villes ou la boutique est présente |
| **street** | [ string ] | Liste des quartiers |
| **state** | [ string ] | Liste des états |
| **country** | [ string ] | Liste des pays |
| **phone** | [ string ] | Liste des contacts à travers lesquels on peut joindre la boutique |

**Remarque :** Grâce à la flexibilité des BD NoSQL , les types des champs ne sont pas des contraintes absolues pour former des documents dans nos collections