
TP- Mongo DB

Objectifs du TP : Se familiariser avec les bases de données NOSQL, orientées documents, avec Mongo DB en partant d'un modèle relationnel.

Exercice 1 :

En se basant sur la description de la base relationnelle des voyageurs séjournant dans des logements (en dessous). Le but est de transformer cette base en une collection de documents JSON.

- Proposez un document JSON représentant *toutes* les informations disponibles sur un des logements, par exemple *U Pinzutu*. On devrait donc y trouver les activités proposées.
- Proposez un document JSON représentant *toutes* les informations disponibles sur un voyageur, par exemple Phileas Fogg.
- Proposez un schéma JSON pour des documents représentant les logements et leurs activités mais pas les séjours.
- Vérifiez la validité syntaxique et insérez les documents dans MongoDB en effectuant une validation avec le schéma.

La base des voyageurs

La base de données « Voyageurs » décrit les pérégrinations de quelques voyageurs plus ou moins célèbres. Ces voyageurs occupent occasionnellement des logements pendant des périodes plus ou moins longues, et y exercent (ou pas) quelques activités.

Voici le schéma de la base. Les clés primaires sont en **gras**, les clés étrangères en *italiques*. Essayez de vous figurer les dépendances fonctionnelles et la manière dont elles permettent de rassembler des informations réparties dans plusieurs tables.

- Voyageur (**idVoyageur**, nom, prénom, ville, région)
- Séjour (**idSéjour**, *idVoyageur*, *codeLogement*, début, fin)
- Logement (**code**, nom, capacité, type, lieu)
- Activité (**codeLogement**, **codeActivité**, description)

La table des voyageurs

La table **Voyageur** ne comprend aucune clé étrangère. Les voyageurs sont identifiés par un numéro séquentiel nommé **idVoyageur**, incrémenté de 10 en 10 (on aurait pu incrémenter de 5, ou de 100, ou changer à chaque fois: la seule chose qui compte est que chaque identifiant soit unique). On indique la ville et la région de résidence.

idVoyageur	nom	prénom	ville	région
10	Fogg	Phileas	Ajaccio	Corse
20	Bouvier	Nicolas	Aurillac	Auvergne
30	David-Néel	Alexandra	Lhassa	Tibet
40	Stevenson	Robert Louis	Vannes	Bretagne

Remarquez que nos régions ne sont pas des régions administratives au sens strict: cette base va nous permettre d'illustrer l'interrogation de bases relationnelles, elle n'a aucune prétention à l'exatitute.

La table Logement

La table **Logement** est également très simple, son schéma ne contient pas de clé étrangère. La clé est un code synthétisant le nom du logement. Voici son contenu.

code	nom	capacité	type	lieu
pi	U Pinzutu	10	Gîte	Corse
ta	Tabriz	34	Hôtel	Bretagne
ca	Causses	45	Auberge	Cévennes
ge	Génépi	134	Hôtel	Alpes

L'information nommée **région** dans la table des voyageurs d'appelle maintenant **lieu** dans la table **Logement**. Ce n'est pas tout à fait cohérent, mais correspond à des situations couramment rencontrées où la même information apparaît sous des noms différents. Nous verrons que le modèle relationnel est équipé pour y faire face.

La table des séjours

Les séjours sont identifiés par un numéro séquentiel incrémenté par unités. Le début et la fin sont des numéros de semaine dans l'année (on fait simple, ce n'est pas une base pour de vrai).

idSéjour	idVoyageur	codeLogement	début	fin
1	10	pi	20	20
2	20	ta	21	22
3	30	ge	2	3
4	20	pi	19	23
5	20	ge	22	24
6	10	pi	10	12
7	30	ca	13	18
8	20	ca	21	22

Séjour contient deux clés étrangères: l'une référençant le logement, l'autre le voyageur. On peut que la valeur de **idVoyageur** (ou **codeLogement**) dans cette relation est *toujours* la valeur de l'une des clés primaire de **Voyageur** (respectivement **Logement**). Si ce n'est pas clair, vous pouvez revoir la définition des clés étrangères et méditer dessus le temps qu'il faudra.

Note

La clé étrangère **codeLogement** n'a pas la même nom que la clé primaire dont elle reprend les valeurs (**code** dans **logement**). Au contraire, **idVoyageur** est aussi bien le nom de la clé primaire (dans **Voyageur**) que de la clé étrangère (dans **Séjour**). Les deux situations sont parfaitement correctes et acceptables. Nous verrons comment spécifier avec SQL le rôle des attributs, indépendamment du nommage.

Connaissant un séjour, je connais donc les valeurs de clé du logement et du voyageur, et je peux trouver la description complète de ces derniers dans leur table respective. ce schéma, comme tous les bons schémas, élimine donc les redondances sans perte d'information.

La table Activité

Cette table contient les activités associées aux logements. La clé est la paire constituée de (codeLogement, codeActivité).

Code Logement	CodeActivité	Description
pi	Voile	Pratique du dériveur et du catamaran
pi	Plongée	Baptêmes et préparation des brevets
ca	Randonnée	Sorties d'une journée en groupe
ge	Ski	Sur piste uniquement
ge	Piscine	Nage loisir non encadrée

Le schéma de cette table a une petite particularité: la clé étrangère `codeLogement` fait partie de la clé primaire. Tout se passe dans ce cas comme si on identifiait les activités relativant au logement auquel elle sont associées. Il s'agit encore une fois d'une situation normale, issue d'un de choix de conception assez courant.

Exercice 2 :

Le service informatique d'une université a décidé de représenter ses données sous forme de documents structurés pour faciliter les processus analytiques. Voici un exemple de documents centrés sur les étudiants et incluant les Unités d'Enseignement (UE) suivies par chacun.

```
{
  "_id": 978,
  "nom": "Jean Dujardin",
  "UE": [{ "id": "ue:11", "titre": "Java", "note": 12 },
        { "id": "ue:27", "titre": "Bases de données", "note": 17 },
        { "id": "ue:37", "titre": "Réseaux", "note": 14 }
  ]
}
{
  "_id": 476,
  "nom": "Vanessa Paradis",
  "UE": [{ "id": "ue:13", "titre": "Méthodologie", "note": 17 },
        { "id": "ue:27", "titre": "Bases de données", "note": 10 },
        { "id": "ue:76", "titre": "Conduite projet", "note": 11 }
  ]
}
```

- Sachant que ces documents sont produits à partir d'une base relationnelle, reconstituez le schéma de cette base et indiquez le contenu des tables correspondant aux documents ci-dessus.
- Proposez une autre représentation des mêmes données, centrée cette fois, non plus sur les étudiants, mais sur les UEs.

Avec les documents semi-structurés, on choisit de privilégier certaines entités, celles qui sont proches de la racine de l'arbre. En centrant sur les UEs, on obtient le même contenu, mais avec une représentation très différente.