

Projet base de données : OrientDB

BONY Audrey

DAVID Oriane

LAGROY DE CROUTTE DE SAINT-MARTIN Anne-Victoire

PAUTREL Léa

ZHAO Junyi

Master 2 Statistiques 2020 Christine Largouet

Table des matières

[I. Présentation du logiciel 1](#_Toc56546027)

[I.1. Objectif et approche 1](#_Toc56546028)

[I.2.Type de données 1](#_Toc56546029)

[I.3 Langage de requête 2](#_Toc56546030)

[I.4. Manuel d’utilisation 2](#_Toc56546031)

[1.4.1 Installation 2](#_Toc56546032)

[1.4.2 Utilisation de OrientDB Studio 4](#_Toc56546033)

[II. Jeu de données 6](#_Toc56546034)

[II.1. Présentation du jeu de données 6](#_Toc56546035)

[II.2. Exemple de requêtes sur le Studio 6](#_Toc56546036)

[III. Python 10](#_Toc56546037)

[Conclusion 11](#_Toc56546038)

# Présentation du logiciel

## I.1. Objectif et approche

OrientDB est un système de gestion de base de données (SGBD) **NoSQL** orienté document-graphe.

Un système NoSQL ne respecte pas le paradigme relationnel basé sur la théorie des ensembles régissant les système SQL. Dans un système SQL, la modélisation de la donnée est structurée par un ensemble de tables et de clefs. La structure est réfléchie en amont et fixe. Au contraire un système NoSQL suit ses propres lois de modélisation. Les données n’ont pas à être homogènes et indexées. Cela rend un système NoSQL beaucoup plus flexible.

OrientDB est construit comme un système orienté graphe tout en incluant des caractéristiques de base orientée document.

Les bases orientée graphe sont obéissent aux lois de la théorie des graphes. Les données sont parcourues sans avoir recours à un index (parcours d’une table) mais en utilisant les arcs (liens) existant entre les nœuds (individus). Le parcours entre nœuds et arcs est très rapide.

Une structure graphe esquive l’utilisation de jointures lors des requêtes qui peuvent être très couteuses surtout lorsqu’on a beaucoup de données.

En ce point, on peut comparer OrientDB à Neo4j. Les interface utilisateurs des deux logiciels sont d’ailleurs très similaires.

OrientDb est aussi orientée document (comme MongoDB). Les données sont stockées de façon semi-structurées. En effet, les données sont dans un format particulier (en JSON souvent) mais la structure des données est libre.

Pour conclure, un système NoSQL orienté document-graphe permet une gestion efficace de données massives et hétérogènes tout en jouissant d’une structure flexible.

## I.2.Type de données

Une base de données orientée graphe est formée de noyaux (vertex) et d’arcs (edge). Un document est stocké dans un noyau et les arcs forment un pointeur d’un nœud à l’autre.

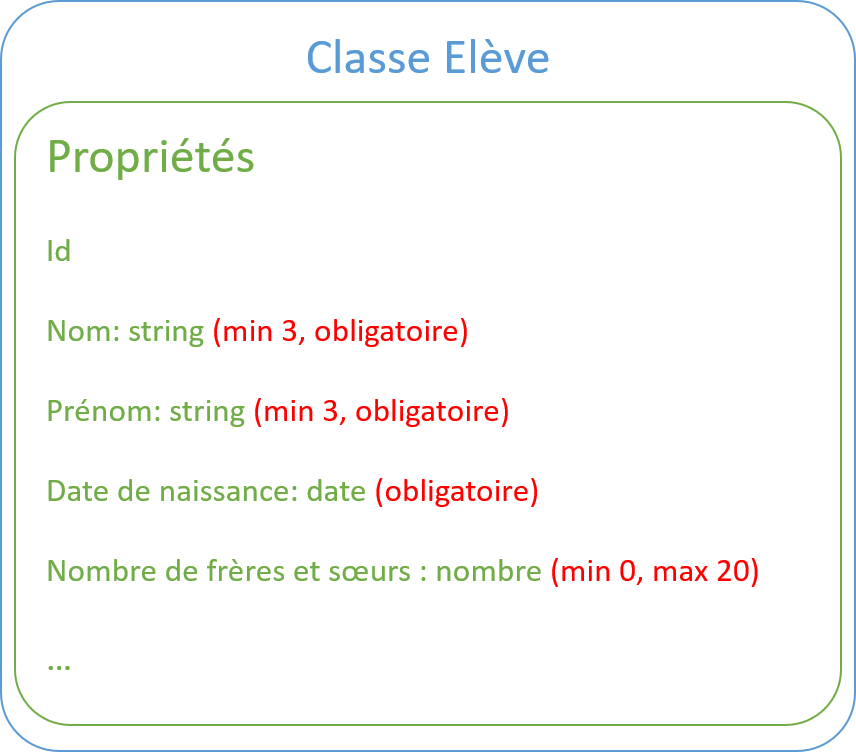
Noyau 1

Noyau 2

Arc

1 relié à 2

Chaque document appartient à une **classe** (ce qui se rapproche de la définition d’une table en système relationnel). Une classe structure et qualifie les attributs des documents. En effet, on pourra lui définir, si on le souhaite, des propriétés puis des contraintes. Une **propriété** définira le nom et le type de l’attribut (chaine de caractère, nombre, etc…). A cela on pourra ajouter des **contraintes** sur les propriétés. On peut par exemple contraindre la forme de l’attribut (nombre de caractères minimum ou maximum, plage de valeurs acceptée) et si l’attribution est obligatoire. En soit les propriétés ne sont pas nécessaires donc un document peut être rempli de façon libre. Cependant un document doit nécessairement avoir un **identifiant**.



CREATE CLASS Eleve

CREATE PROPERTY Eleve.nom STRING

ALTER PROPERTY Eleve.nom MIN 3

ALTER PROPERTY Eleve.nom MANDATORY true

Dans le model, on définira des **super-classes.** Celles-ci indiquent la nature de l’objet : nœud (vertex) ou arc (edge).

## I.3 Langage de requête

Le langage de requête du logiciel OrientDB est en langage SQL étendu afin de répondre aux exigences d’une structure graphe. Ainsi des commandes permettant de naviguer au sein du graphe. Ce langage est relativement basique, ce qui permet à des utilisateurs assez peu expérimentés de pouvoir explorer et utiliser la base de données.

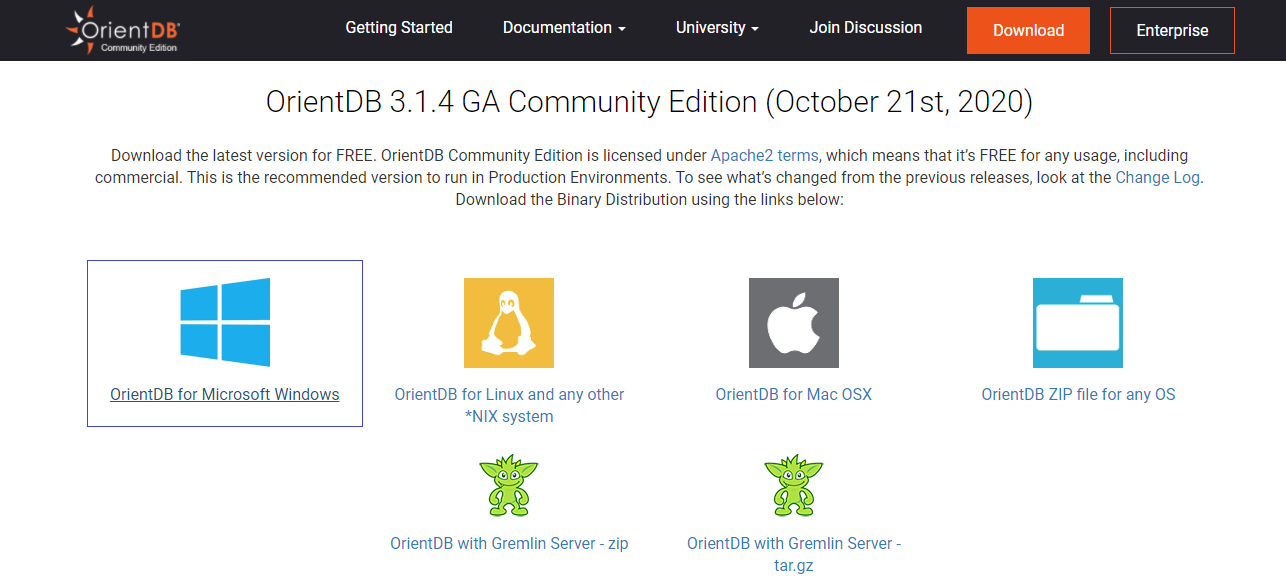
## I.4. Manuel d’utilisation

### 1.4.1 Installation

Pour utiliser la SGBD OrientDB, il faut d’abord l’installer sur son ordinateur.

**Installation d’OrientDB 3.1.4 sur Windows**

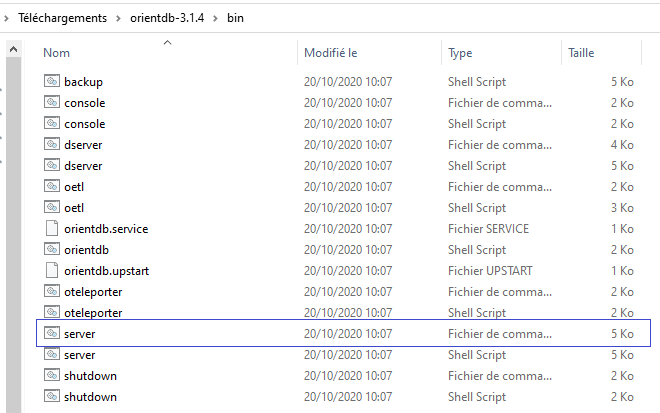
Pour l’installer, rendez-vous sur le site https://www.orientdb.org/download et cliquez sur ‘OrientDB for Microsoft Windows’.



Un fichier ZIP se télécharge alors. A la fin du téléchargement, il suffit de dézipper le dossier et d’en extraire les composantes.

**Configuration d’OrientDB 3.1.4**

Dans le dossier ‘orientdb-3.1.4’, allez dans le dossier ‘bin’. Ouvrez alors fichier de commande ‘server’. **ATTENTION**, il y a 2 fichiers appelés ‘server’, choisissez bien le fichier de commande (extension .bat) si vous avez un système Windows et le Shell Script (extension .sh) si vous avez un système basé sur Unix ( un mac par exemple).

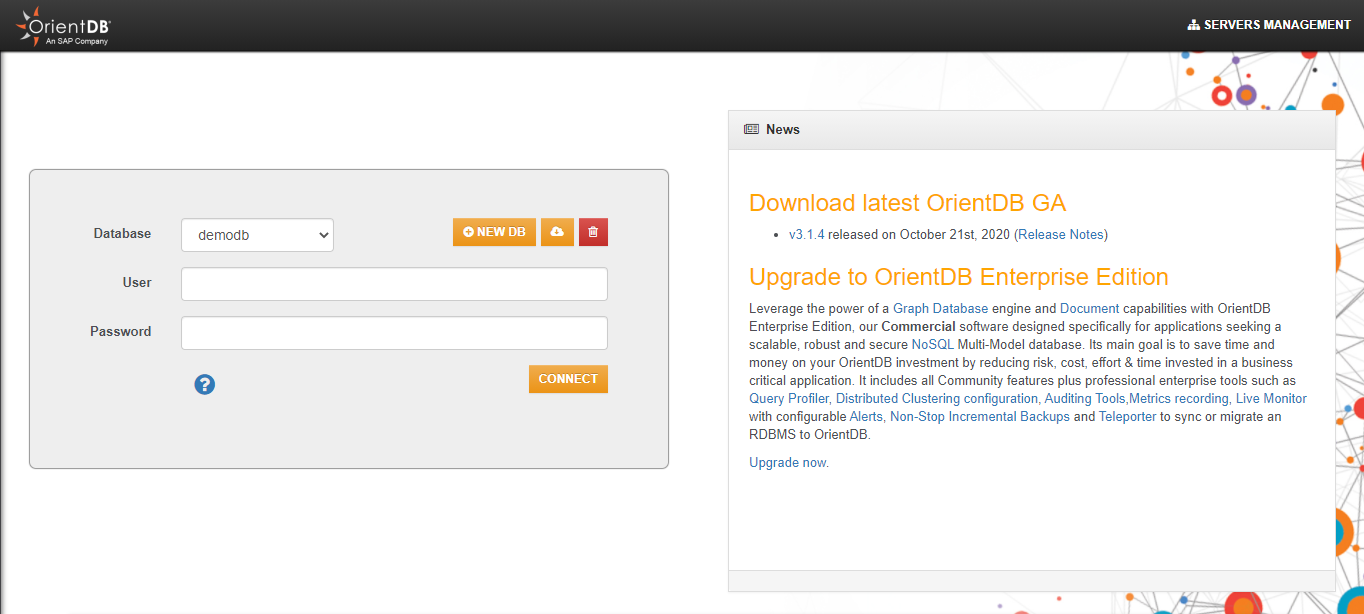


Lorsque vous ouvrez le fichier, on vous affectera par défaut le nom d’utilisateur ‘Root’ et on vous demandera un mot de passe. Vous pouvez donc saisir un mot de passe. **ATTENTION**, votre mot de passe sera à réutiliser donc notez-le bien ! De plus, ne faites pas ‘Enter’ sans en saisir sinon cela posera un problème par la suite. En effet, le mot de passe choisi sera aléatoire et il sera difficile de le récupérer.

Si la fenêtre du serveur se referme directement sans que vous puissiez configurer, c’est que votre version de Java n’est pas à jour. Vous pouvez télécharger la dernière version sur ce lien : <https://www.java.com/fr/download/>.

**Ouverture d’OrientDB Studio**

Une fois la configuration du nom d’utilisateur et mot de passe effectuée, la fenêtre d’OrientDB Studio devrait s’ouvrir automatiquement dans votre navigateur. Si elle ne s’ouvre pas, vous pouvez l’ouvrir avec ce lien : <http://localhost:2480/studio/index.html#/> après avoir quand même ouvert le server OrientDB. Vous arrivez alors à la page d’accueil suivante :

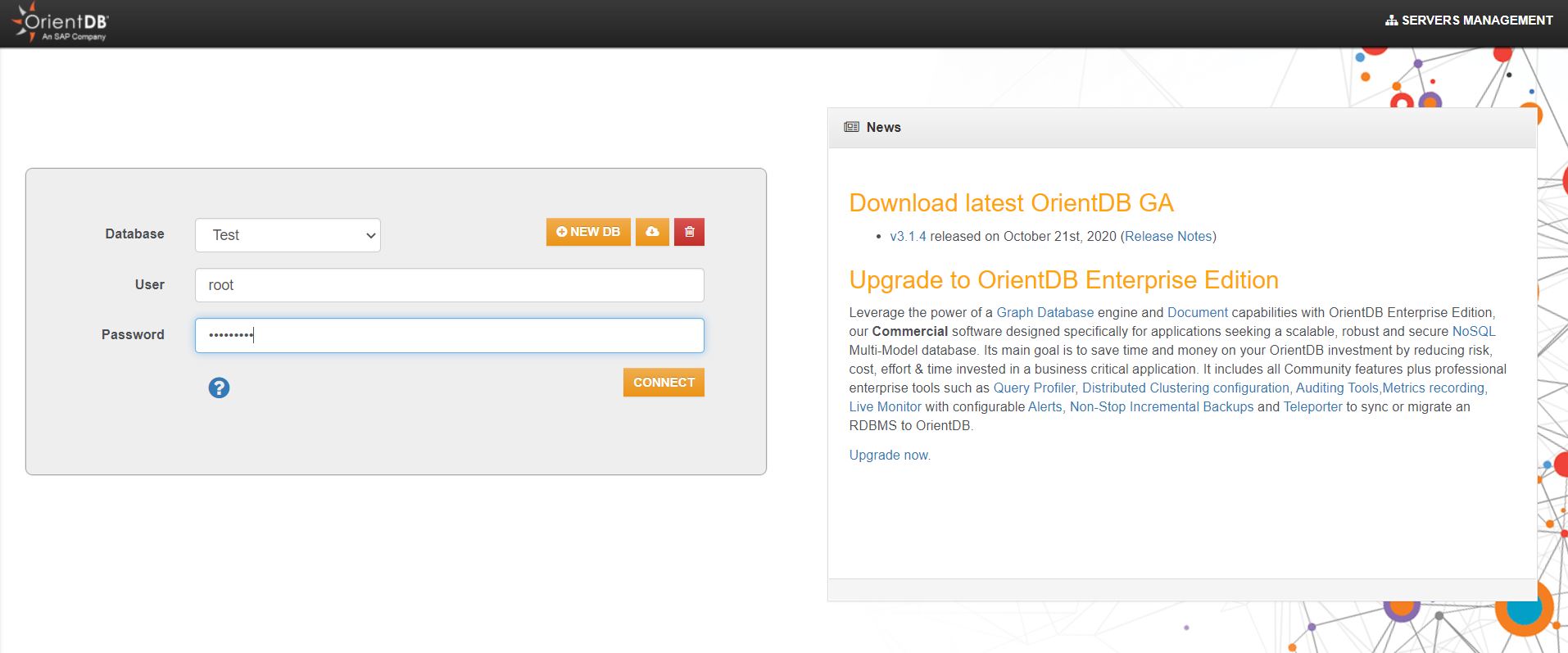


**ATTENTION :** Si vous vous connectez pour la seconde fois, il faut rouvrir le fichier serveur (avec l’extension .bat ou .sh selon votre système d’exploitation). Vous pourrez ensuite vous connecter au serveur avec le lien.

### 1.4.2 Utilisation de OrientDB Studio

Un fois sur la page d’accueil de OrientDB Studio, vous pouvez choisir la base de données sur laquelle vous voulez travailler (1), indiquer l’utilisateur « root » (2) et renseigner votre mot de passe (3). Votre mot de passe est celui que vous avez choisi lors de la configuration d’OrientDB.

Vous pouvez aussi créer une nouvelle base de données (4), importer une base de données publiques qui existe déjà (5) ou supprimer la base de données sélectionnée (6).



1

2

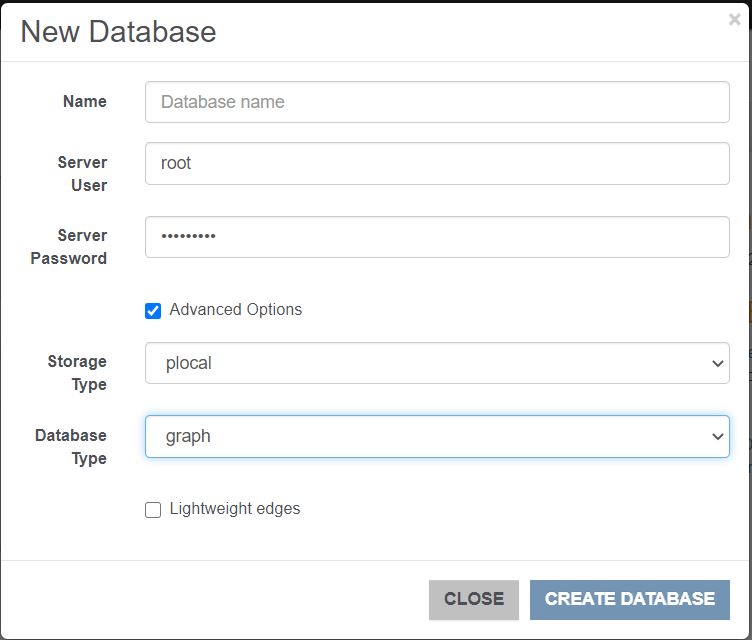
3

4

5

6

Lorsque vous créez une base de données, vous pourrez lui donner un nom (7), définir son type (document ou graph) (8) ainsi que le mot de passe associé à la base de données (9).



7

8

9

Dans OrientDB, vous avez le choix d’effectuer une action de manière visuelle ou à l’aide d’une ligne de code. L’interface visuelle de manipulation de la base est dans l’onglet *Schema* et l’interface permettant l’utilisation de ligne de code est dans l’onglet *Browse*. Nous allons à présent détailler quelques actions simples pour modifier votre base de données à l’aide d’OrientDB Studio.

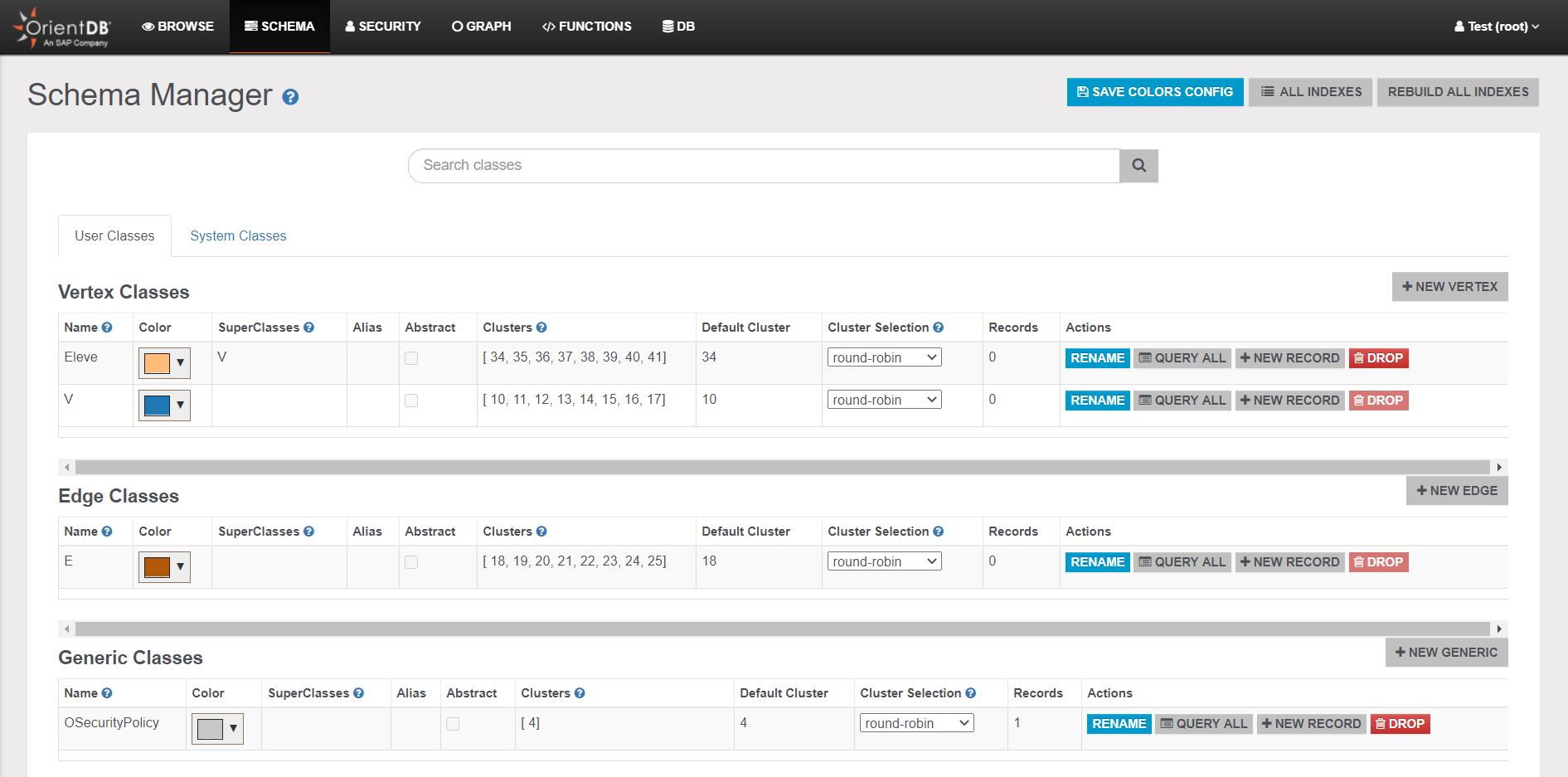
***Ajouter, modifier, supprimer une classe***

Pour créer une classe de type nœuds, on peut appuyer sur le bouton + *New Vertex* (1). On aura alors juste à indiquer le nom de la classe qu’on souhaite créer.

La commande équivalente à ces actions est :

CREATE CLASS Eleve extends V

On peut renommer la classe (2), changer la couleur d’affichage de la classe (3), afficher l’ensemble des documents enregistrés dans la classe (4) (le nombre d’enregistrement est affiché dans la colonne *Records*), ajouter un document (5) ou supprimer la classe (6).



1

2

3

4

5

6

Pour résumer :

|  |  |
| --- | --- |
| Action | Commande Browse |
| 1 | CREATE CLASS Eleve extends V |
| 2 | ALTER CLASS Eleve NAME Eleve2 |
| 4 | SELECT \* FROM Eleve |
| 5 | INSERT INTO Eleve SET nom = … , prenom = … |
| 6 | DROP CLASS Eleve |

***Ajouter, modifier, supprimer la propriété d’une classe***

# Jeu de données

## II.1. Présentation du jeu de données

Le jeu provient de la base jeu de données publiques du logiciel OrientDB.

Il s’appelle 'Tolkien-Arda'. Il décrit l’univers des romans de Tolkien (Trilogie du *Seigneur des Anneaux* et *Le Hobbit*)

Database containing Tolkien (Middle-earth) related information like Characters, Locations and Events from his different works (<http://arda-maps.org>)

## II.2. Exemple de requêtes sur le Studio

Le langage de requête est le SQL. Certains éléments de langage ont en effet été ajouter afin de pouvoir mieux manipuler les relations entre «Vertex ». Dans le Studio, les requêtes peuvent se faire dans deux onglets distincts selon le type de sortie que l’on veut avoir : « graph » (pour avoir une sortie graph) et « browse » pour un tableau de données.

* Principes simples requêtes SQL :

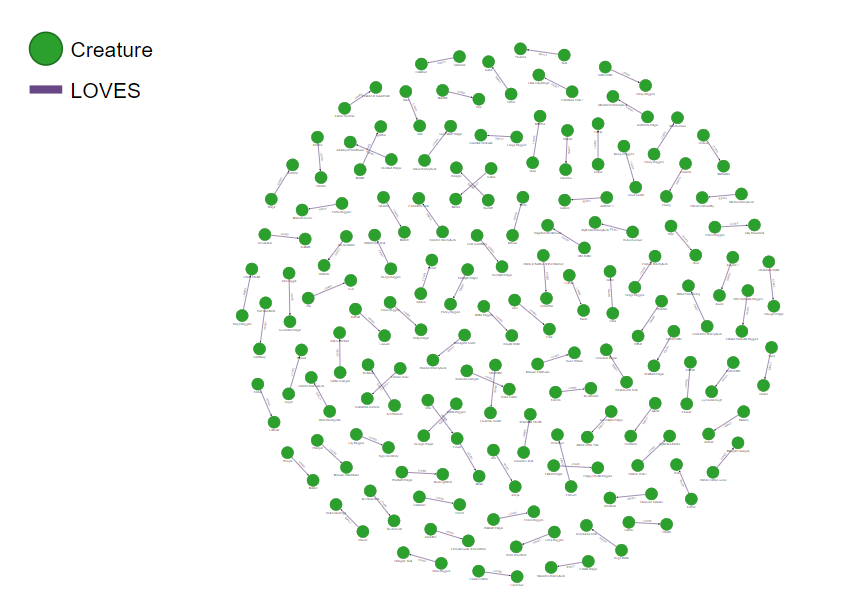
Les requêtes SQL de base s’appliquent. Les relations ou « edges » sont considérées comme une simple table ou l’on peut sélectionner des éléments.

* *Qui aime qui?*

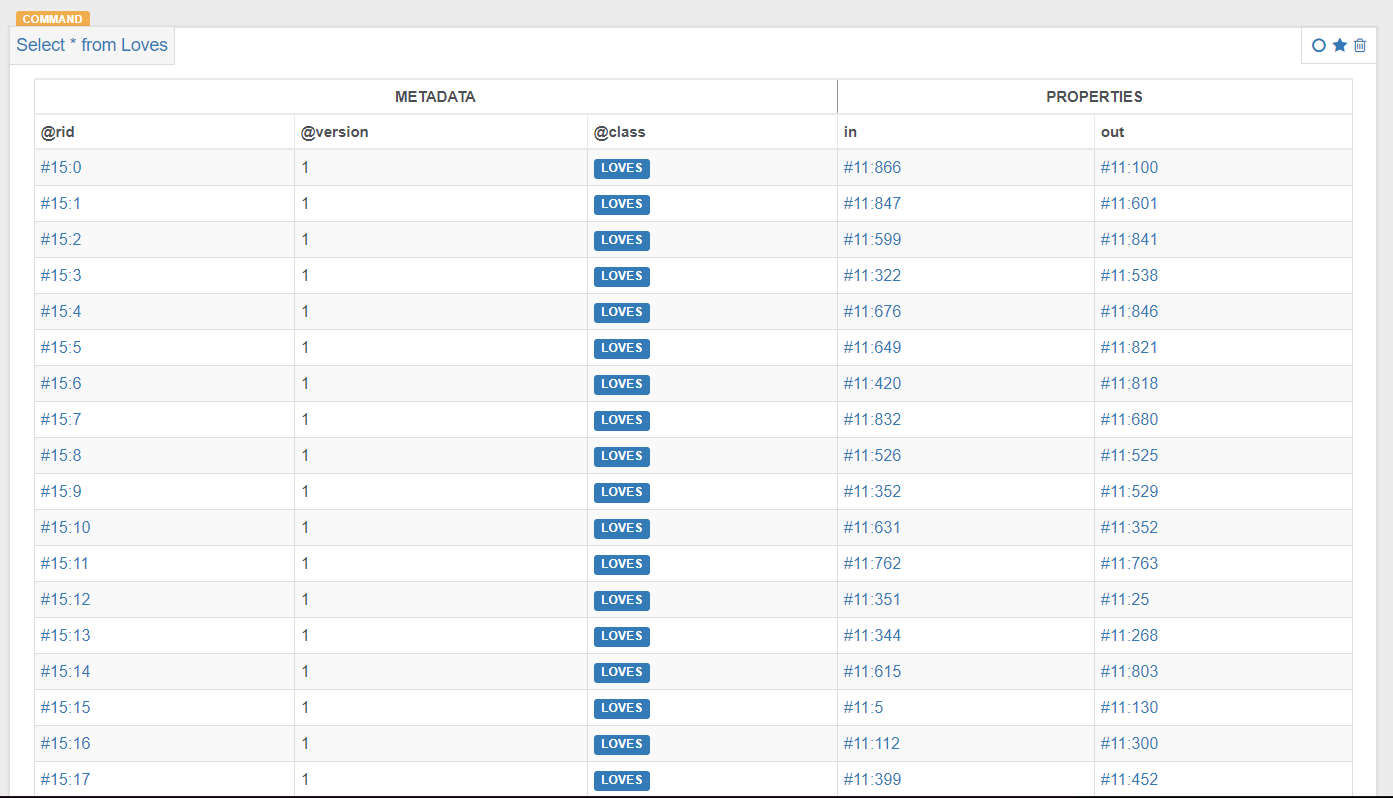
select \* from Loves

Sortie graphe :

On observe les « vertex » reliés par les « edges » lorsque l’on recherche le « edges »



Sortie Tableau de données :



On observe les caractéristiques des relations : élément 1 de la relation, élément 2 (in et out).

* Agrégations :

On peut également faire des agrégations : comme en SQL normal, il est possible d’appliquer les fonctions mathématiques et les fonctions de regroupement dans Group by. On s’attend ici à des sorties tableaux, ces commandes sont généralement écrite dans la console de l’onglet « browse ».

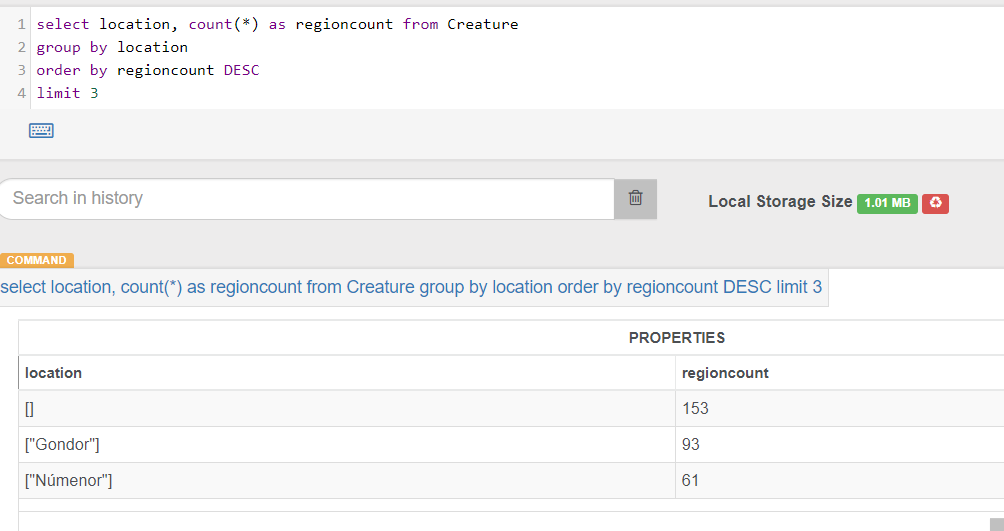
* *Quelles sont les trois régions les plus peuplées?*

select location, count(\*) as regioncount from Creature

group by location

order by regioncount DESC

limit 3



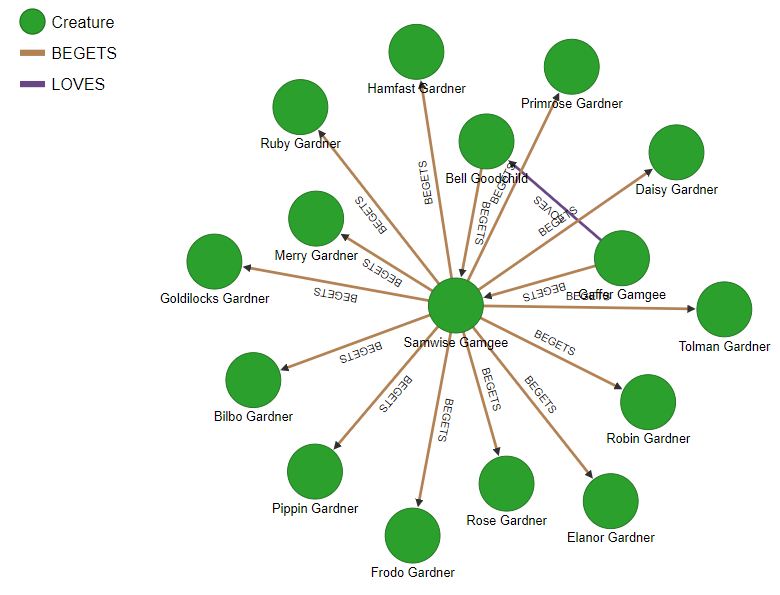
* Relation *« edges »* entre éléments *« vertex »*

Notion de distance/chemin: Les relations peuvent être désignées par le nom MATCH. L’ajout de $Pathelements permet d‘afficher le graphe relatif à tous les éléments que l’on a désigné ainsi que leur relation. La flèche permet de polariser la relation. Ainsi le résultat à la requête suivante est faux :

* *Qui sont les enfants de Sam Gangee ?*

MATCH {Class: Creature, as: Father, where: (name='Samwise Gamgee')}-BEGETS-{Class: Creature, as: Children}

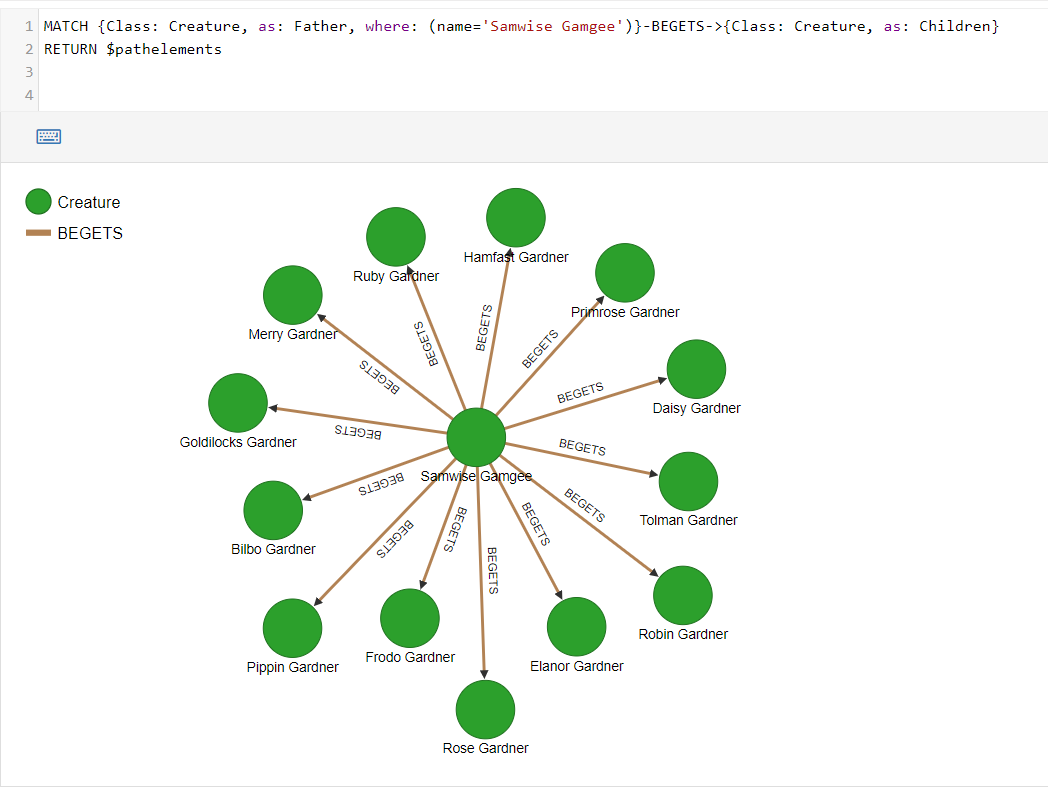
RETURN $pathelements



Il faut polariser la relation, afin d’avoir uniquement les Creatures que Samwise a engendré et ne pas avoir celle qui ont engendré Samwise (ses propres parents).

MATCH {Class: Creature, as: Father, where: (name='Samwise Gamgee')}-BEGETS-> {Class: Creature, as: Children}

RETURN $pathelements



* *Combien d’enfants a Samwise ? On attend une sortie tableau…*

Il suffit de modifier le return par :

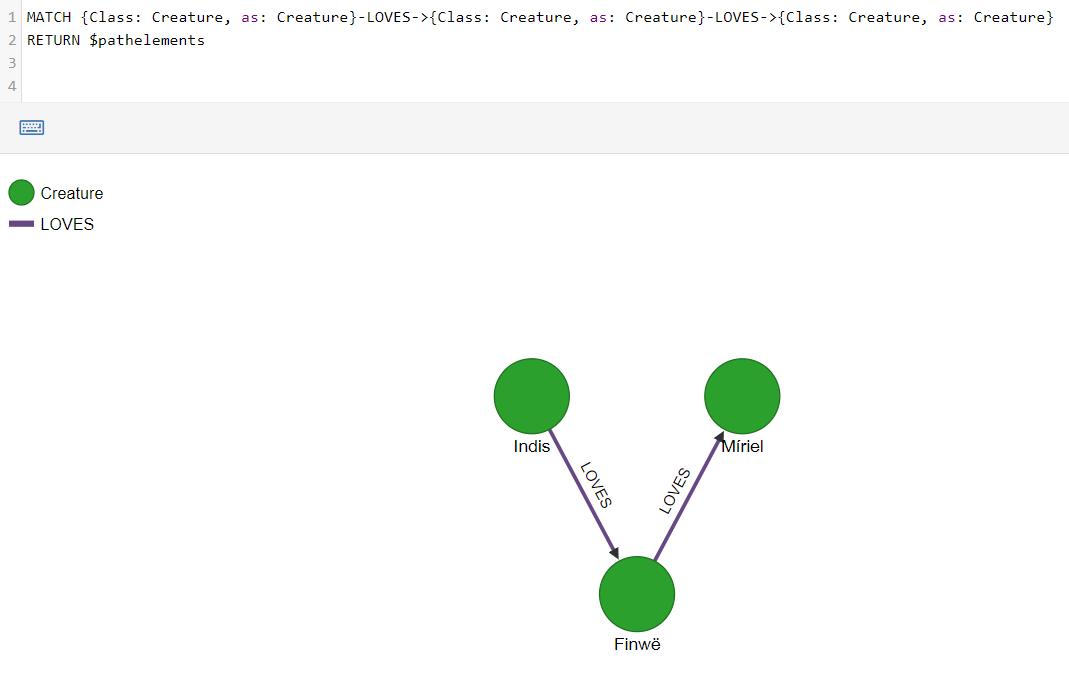
RETURN name, count(Children)

On obtient la réponse 13.

* *Existe-t-il un triangle amoureux?*

*MATCH {Class: Creature, as: Creature}-LOVES->{Class: Creature, as: Creature}-LOVES->{Class: Creature, as: Creature}*

*RETURN $pathelements*



* Notion de chemin

La requête suivante illustre la notion de chemin. En effet, il existe l’élément Path permettent de relier par un chemin les éléments désignées d’un graphe.

🡪*Comment sont reliés Bilbo Baggins et Frodo Baggins ?*

SELECT $path as path

FROM (TRAVERSE outE(), inV() FROM (SELECT @rid FROM Creature WHERE name='Aragorn II'))

WHERE @rid==(SELECT @rid FROM Creature WHERE name='Isildur')

La fonction shortestPath permet de désigner le chemin le plus court.

* *Quels sont les ancêtres qui séparent Aragorn II et Isildur ?*

SELECT expand(path) FROM (

SELECT shortestPath($from, $to) AS path

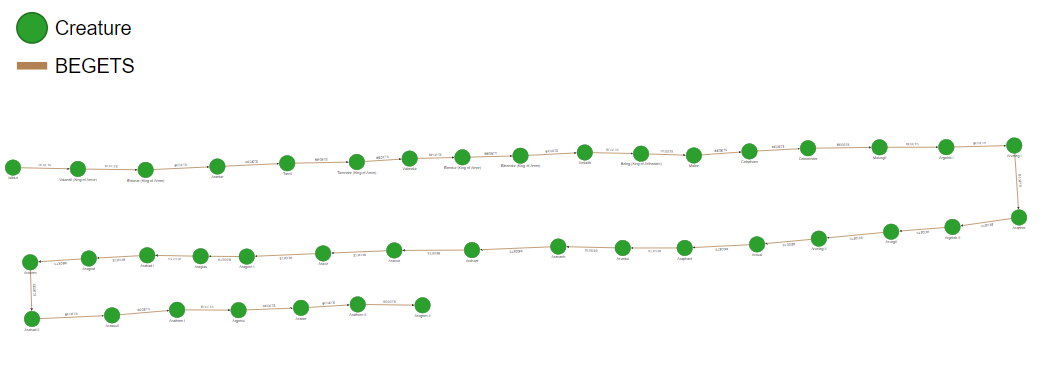
LET

$from = (SELECT FROM Creature WHERE name='Aragorn II'),

$to = (SELECT FROM Creature WHERE name='Isildur')

UNWIND path

)



* *Combien de générations les séparent (en comptant les leurs) ? On attend une sortie tableau…*

Il suffit de rajouter un count (\*) à partir de la requête précédente. On obtient un résultat égal à 40.

# Python

Il existe un package de Python permettant l’utilisation de OrientDB directement sur Python. Ce package est extrêmement complexe et requiert de bonnes connaissances de programmation. Malgré de nombreuses heures de recherche passée à découvrir le package nous avons décidé de ne pas l’utiliser. En effet, le package présente deux modules : « Client » et « OGM ». Le module « Client est le module de « base ». Il permet la réalisation de taches simple : se connecter à un serveur, manipuler les bases de données et faire des requêtes SQL. Il y a plusieurs reproches que nous pouvons faire à PyOrient. Tout d’abord, il est très compliqué de se connecter à un serveur. En effet, il existe un problème de version du package rendant très compliqué la connexion. Le « bug » est reporté à de nombreuses reprises sur le web. Nous avons tenté plusieurs stratégies de résolution du bug. Malgré tout, aucune n’a été 100% efficace sur l’ensemble de nos machines.

La deuxième difficulté que présente PyOrient est la non-lisibilité du résultat des requêtes sur Python directement. Une fois le résultat stocké dans une variable, il est nécessaire d’utiliser des outils de code comme Flask ou Django pour lire les résultats. Cela nous semblait ambitieux et de plus, pour un résultat incertain car nous n’aurions certainement pas obtenus de graphes mais des lignes de résultats ?. L’utilisation de l’ORM est encore plus complexe car réservée aux bases de taille importante et aux grands graphes.

# Conclusion

Documentation mal faite (pas constant, info manquante)

Langage utilisé basé SQL donc assez facile à prendre en main pour les personnes maitrisant des rudiments en base de données