Neo4j – TP
SNCF MAX

Asencio Victor, Brunet Maxence, Peze Thomas

1. Choix du modèle relationnel

La quantité d'informations a traiter est importante. Il donc important de s'accorder sur le modèle relationnel le plus efficace possible. Notre modèle sera le suivant :

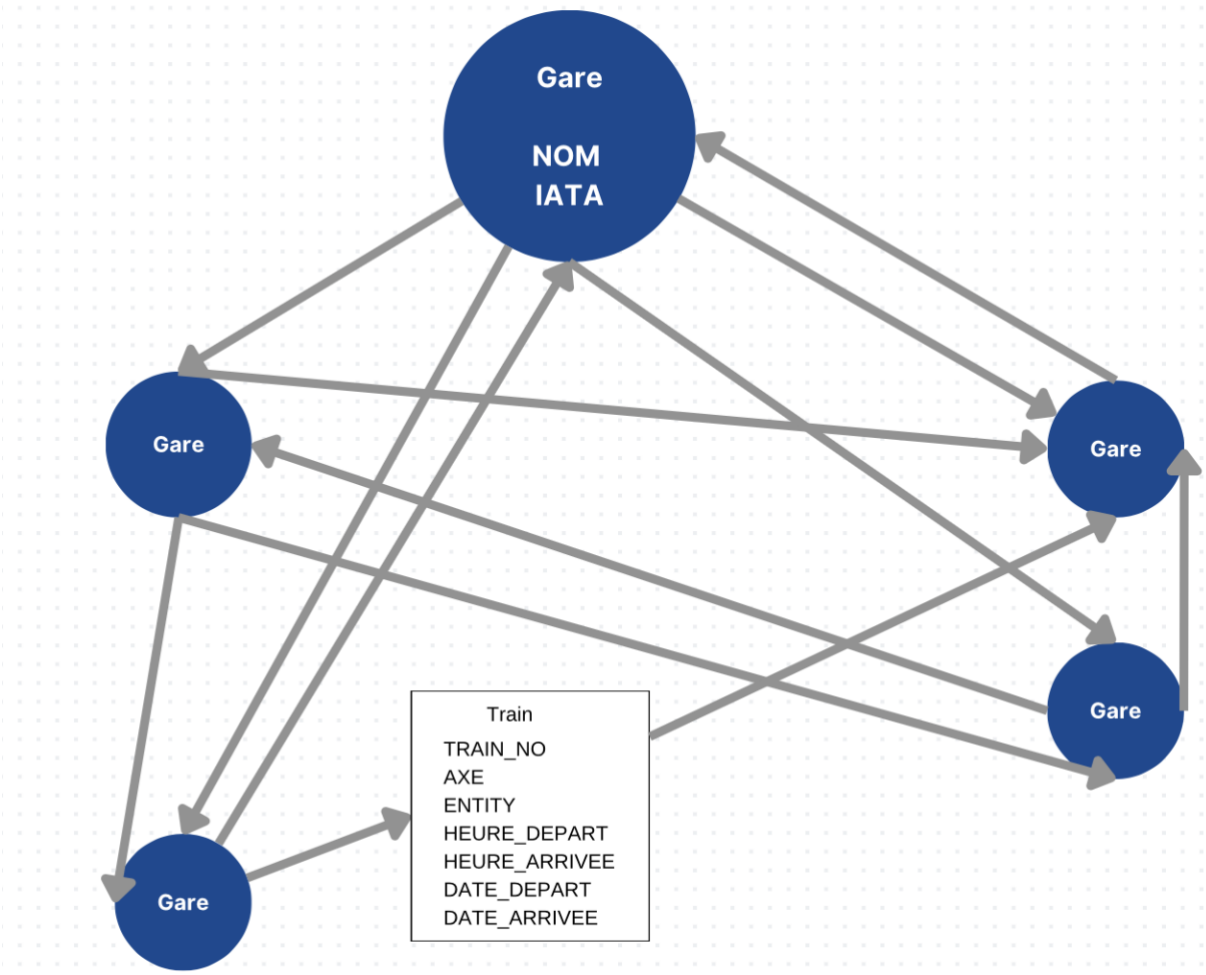


Figure 1: Modèle relationnel

Nous avons établi ce modèle relationnel après avoir identifié les problèmes liés à la date et aux horaires de voyage des trains. Pour limiter les conflits liés aux trains circulant sur deux jours (trains de nuit) nous avons créé une nouvelle colonne « Date arrivée » dans le fichier. Si le train voyage sur deux jours alors Date-arrivée : (Date_départ) J+1. Sachant que Date_départ = Date.

2. Nettoyage du dataset - Nettoyage Data.py

Après une rapide étude du dataset, on remarque la présence de lignes inexploitables (case vide ou « TBD »). On supprime toutes ces cases ainsi que la colonne « Disponibilité de places MAX JEUNE et MAX SENIOR ».
Notre dataset est désormais exploitable. On notera qu'il y a 302 gares différentes dans le fichier, il faudra donc obtenir 302 nœuds dans notre base de données.

3. Création des nœuds – Nœuds Gare.py

Avec la commande MERGE on peut ajouter les nœuds à notre base de données neo4j. Cependant, en utilisant la méthode vue en cours cette opération est très longue. On utilise donc une autre méthode plus rapide : ThreadPoolExecutor().

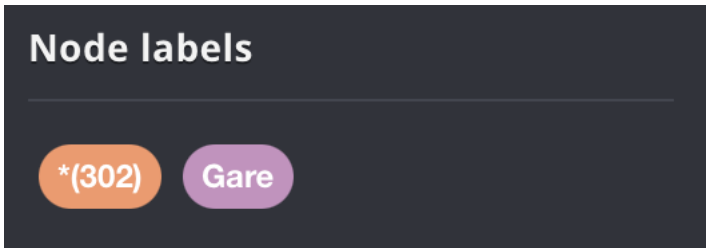


Figure 2: Nombre total de nœuds après exécution du programme

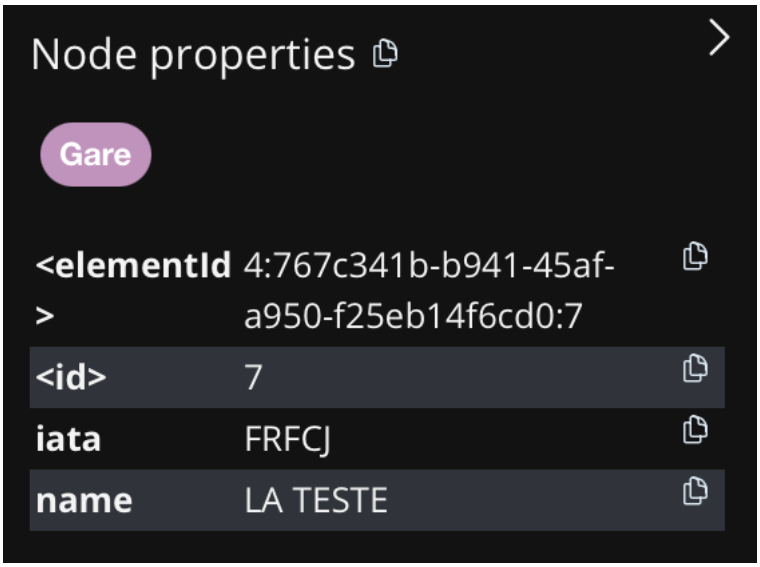


Figure 3: Structure et attributs d'un nœud

Nos nœuds ont pour attributs un nom et un IATA. Chaque nœud représente donc une gare unique (pas de doublons).

4. Ajout des liaisons - Liaisons Train.py

Pour ajouter les liaisons nous utilisons de nouveau ThreadPoolExecutor(). Même avec cette méthode, l'exécution de ce programme a pris plus de quatre heures. Chaque liaison est représentée par un trajet, effectué par un train. Chaque trajet a donc un numéro de train, un axe, une entité, une heure de départ et d'arrivée ainsi qu'une date de départ et d'arrivée. L'ajout des liaisons nous a pris plus de 4 heures. Au final on obtient tous les trajets uniques de notre fichier source.

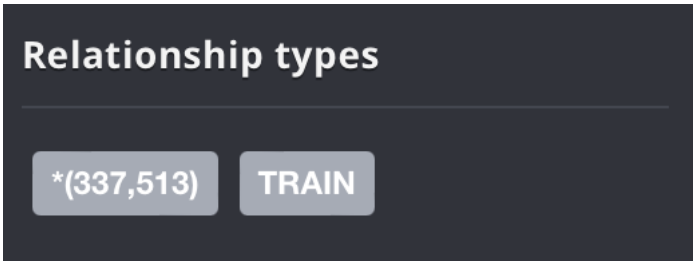


Figure 4: Nombre total de relations

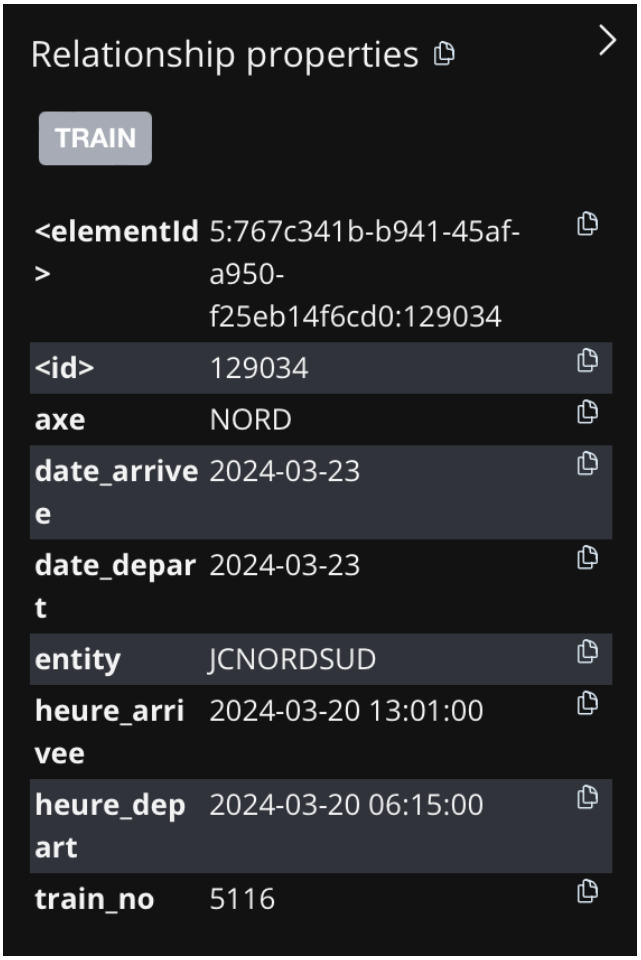


Figure 5: Structure d'une liaison entre deux nœuds

5. Recherche dans la base de données – Requêtes_1.py

Pour rechercher les trajets dans notre base, on utilise une commande MATCH avec des liaisons pour trouver les trajets possibles (ou non). On est capable d'afficher des trajets directs et indirects avec une ou plusieurs correspondances. Le seul problème que nous n'avons pas adressé dans ce code, c'est la gestion des entrées utilisateurs. Il faut écrire les noms de gares comme ils sont écrits dans la BDD (Bordeaux ne marchera pas, il faut rentrer BORDEAUX ST JEAN). Notre code est sensible à la casse.

```
In [4]: runfile('/Users/victorasencio/Desktop/ESME/Neo4j/Requêtes_1.py', wdir='/Users/
victorasencio/Desktop/ESME/Neo4j')
Ville de départ : BANYULS SUR MER
Ville d'arrivée : GAP
Trajets directs :
Aucun trajet direct trouvé.

Trajets indirects avec correspondance :
- BANYULS SUR MER -> ['PARIS (intramuros)'] -> GAP
- BANYULS SUR MER -> ['PARIS (intramuros)'] -> GAP
- BANYULS SUR MER -> ['PARIS (intramuros)'] -> GAP
- BANYULS SUR MER -> ['PARIS (intramuros)'] -> GAP
- BANYULS SUR MER -> ['PARIS (intramuros)'] -> GAP
- BANYULS SUR MER -> ['PARIS (intramuros)'] -> GAP
- BANYULS SUR MER -> ['PARIS (intramuros)'] -> GAP
- BANYULS SUR MER -> ['PARIS (intramuros)'] -> GAP
- BANYULS SUR MER -> ['PARIS (intramuros)'] -> GAP
- BANYULS SUR MER -> ['PARIS (intramuros)'] -> GAP
- BANYULS SUR MER -> ['PARIS (intramuros)'] -> GAP
- BANYULS SUR MER -> ['PARIS (intramuros)'] -> GAP
- BANYULS SUR MER -> ['PARIS (intramuros)'] -> GAP
- BANYULS SUR MER -> ['PARIS (intramuros)'] -> GAP
- BANYULS SUR MER -> ['PARIS (intramuros)'] -> GAP
- BANYULS SUR MER -> ['PARIS (intramuros)'] -> GAP
- BANYULS SUR MER -> ['PARIS (intramuros)'] -> GAP
```

Figure 6: Exemple de trajet réalisable en une correspondance

```
Console 1/A
- NICE VILLE -> ['MARNE LA VALLEE CHESSY', 'BORDEAUX ST JEAN'] -> LA TESTE
- NICE VILLE -> ['MARNE LA VALLEE CHESSY', 'BORDEAUX ST JEAN'] -> LA TESTE
- NICE VILLE -> ['MARNE LA VALLEE CHESSY', 'BORDEAUX ST JEAN'] -> LA TESTE
- NICE VILLE -> ['MARNE LA VALLEE CHESSY', 'BORDEAUX ST JEAN'] -> LA TESTE
- NICE VILLE -> ['MARNE LA VALLEE CHESSY', 'BORDEAUX ST JEAN'] -> LA TESTE
- NICE VILLE -> ['MARNE LA VALLEE CHESSY', 'BORDEAUX ST JEAN'] -> LA TESTE
- NICE VILLE -> ['MARNE LA VALLEE CHESSY', 'BORDEAUX ST JEAN'] -> LA TESTE
- NICE VILLE -> ['MARNE LA VALLEE CHESSY', 'BORDEAUX ST JEAN'] -> LA TESTE
- NICE VILLE -> ['MARNE LA VALLEE CHESSY', 'BORDEAUX ST JEAN'] -> LA TESTE
- NICE VILLE -> ['MARNE LA VALLEE CHESSY', 'BORDEAUX ST JEAN'] -> LA TESTE
- NICE VILLE -> ['MARNE LA VALLEE CHESSY', 'BORDEAUX ST JEAN'] -> LA TESTE
- NICE VILLE -> ['MARNE LA VALLEE CHESSY', 'BORDEAUX ST JEAN'] -> LA TESTE
- NICE VILLE -> ['MARNE LA VALLEE CHESSY', 'BORDEAUX ST JEAN'] -> LA TESTE
- NICE VILLE -> ['MARNE LA VALLEE CHESSY', 'BORDEAUX ST JEAN'] -> LA TESTE
- NICE VILLE -> ['MARNE LA VALLEE CHESSY', 'BORDEAUX ST JEAN'] -> LA TESTE
- NICE VILLE -> ['MARNE LA VALLEE CHESSY', 'BORDEAUX ST JEAN'] -> LA TESTE
- NICE VILLE -> ['MARNE LA VALLEE CHESSY', 'BORDEAUX ST JEAN'] -> LA TESTE
- NICE VILLE -> ['MARNE LA VALLEE CHESSY', 'BORDEAUX ST JEAN'] -> LA TESTE
- NICE VILLE -> ['MARNE LA VALLEE CHESSY', 'BORDEAUX ST JEAN'] -> LA TESTE
- NICE VILLE -> ['MARNE LA VALLEE CHESSY', 'BORDEAUX ST JEAN'] -> LA TESTE
- NICE VILLE -> ['MARNE LA VALLEE CHESSY', 'BORDEAUX ST JEAN'] -> LA TESTE
- NICE VILLE -> ['MARNE LA VALLEE CHESSY', 'BORDEAUX ST JEAN'] -> LA TESTE
- NICE VILLE -> ['MARNE LA VALLEE CHESSY', 'BORDEAUX ST JEAN'] -> LA TESTE
- NICE VILLE -> ['MARNE LA VALLEE CHESSY', 'BORDEAUX ST JEAN'] -> LA TESTE

In [35]:
```

Figure 7: Exemple d'un trajet avec plusieurs correspondances

```
In [3]: runfile('/Users/victorasencio/Desktop/ESME/Neo4j/Requêtes_1.py', wdir='/Users/victorasencio/Desktop/ESME/Neo4j')
Ville de départ : BORDEAUX ST JEAN
Ville d'arrivée : DAX
Trajets directs :
Aucun trajet direct trouvé.

Trajets indirects avec correspondance :
Aucun trajet indirect trouvé.
```

Figure 8: Exemple de trajet impossible en tgv

6. Recherche dans la base de données – Requêtes_2.py

Ici on se concentre sur l'implémentation d'une sélection plus raffinée avec l'ajout des horaires de départ et d'arrivée. On obtient toutes les liaisons qui partent avant 15h ou arrivent avant 23h00.

```
PS C:\Users\thpez> & D:/Logiciels/Anaconda/python.exe "d:/OneDrive - ESME/ESME/CSV NEO4J/essaie22.03gpt.py"
Ville de départ : MONTPELLIER SAINT-ROCH
Ville d'arrivée : MARNE LA VALLEE CHESSY
Heure de départ souhaitée (format HH:MM) : 15:00
Heure d'arrivée souhaitée (format HH:MM) : 23:00
Trajets directs :
- MONTPELLIER SAINT-ROCH (2024-03-20 11:59:00) -> MARNE LA VALLEE CHESSY (2024-03-20 15:46:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 08:59:00) -> MARNE LA VALLEE CHESSY (2024-03-20 12:47:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 05:31:00) -> MARNE LA VALLEE CHESSY (2024-03-20 09:17:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 11:59:00) -> MARNE LA VALLEE CHESSY (2024-03-20 15:41:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 05:32:00) -> MARNE LA VALLEE CHESSY (2024-03-20 09:17:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 05:28:00) -> MARNE LA VALLEE CHESSY (2024-03-20 09:17:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 14:51:00) -> MARNE LA VALLEE CHESSY (2024-03-20 18:45:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 14:51:00) -> MARNE LA VALLEE CHESSY (2024-03-20 18:47:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 06:31:00) -> MARNE LA VALLEE CHESSY (2024-03-20 10:24:00)
```

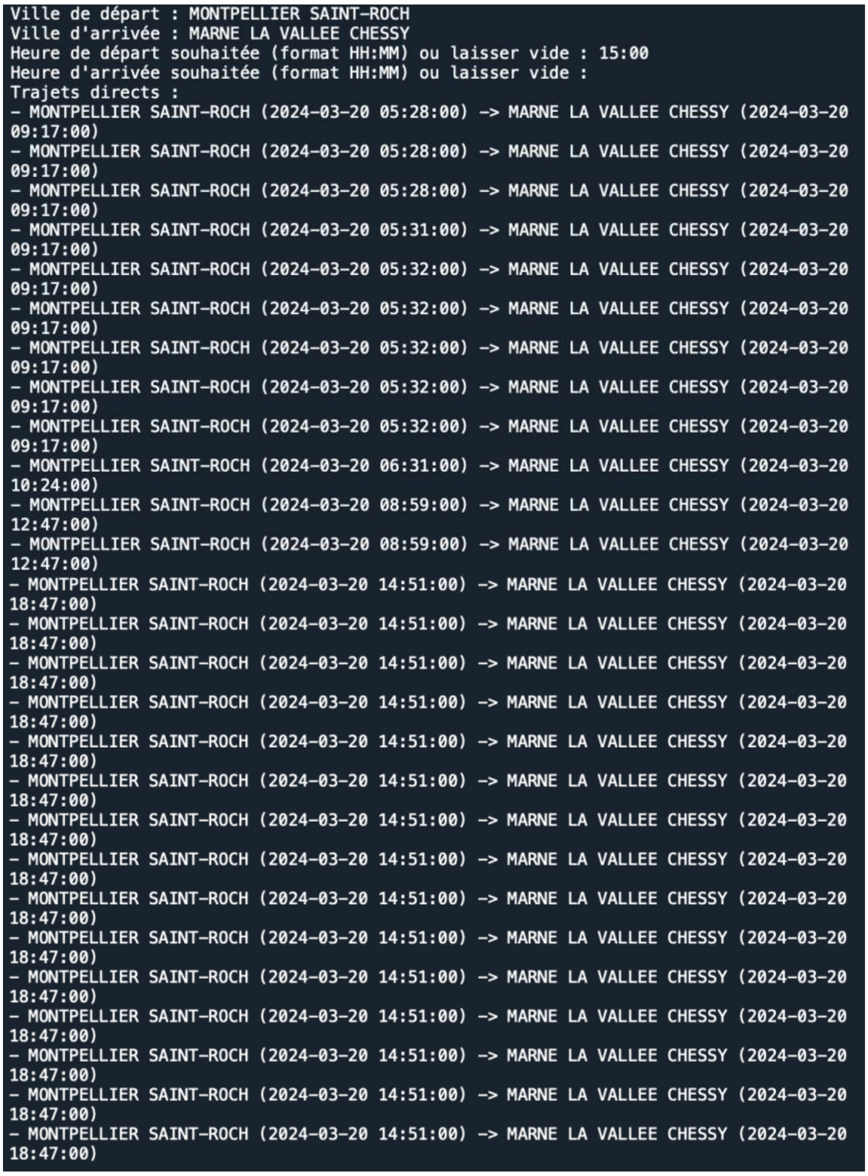
Figure 9: Affichage des trajets avec horaire non trié

Donc ce n'est pas bon. On modifie notre code et on obtient enfin un résultat trié qui affiche les trains qui partent avant 15h00 et qui arrivent aux alentours de 18h00.

```
Ville de départ : MONTPELLIER SAINT-ROCH
Ville d'arrivée : MARNE LA VALLEE CHESSY
Heure de départ souhaitée (format HH:MM) ou laisser vide :
Heure d'arrivée souhaitée (format HH:MM) ou laisser vide : 18:00
Trajets directs :
- MONTPELLIER SAINT-ROCH (2024-03-20 14:51:00) -> MARNE LA VALLEE CHESSY (2024-03-20 18:47:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 14:51:00) -> MARNE LA VALLEE CHESSY (2024-03-20 18:47:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 14:51:00) -> MARNE LA VALLEE CHESSY (2024-03-20 18:47:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 14:51:00) -> MARNE LA VALLEE CHESSY (2024-03-20 18:47:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 14:51:00) -> MARNE LA VALLEE CHESSY (2024-03-20 18:47:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 14:51:00) -> MARNE LA VALLEE CHESSY (2024-03-20 18:47:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 14:51:00) -> MARNE LA VALLEE CHESSY (2024-03-20 18:47:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 14:51:00) -> MARNE LA VALLEE CHESSY (2024-03-20 18:47:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 14:51:00) -> MARNE LA VALLEE CHESSY (2024-03-20 18:47:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 14:51:00) -> MARNE LA VALLEE CHESSY (2024-03-20 18:47:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 14:51:00) -> MARNE LA VALLEE CHESSY (2024-03-20 18:47:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 14:51:00) -> MARNE LA VALLEE CHESSY (2024-03-20 18:47:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 14:51:00) -> MARNE LA VALLEE CHESSY (2024-03-20 18:47:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 14:51:00) -> MARNE LA VALLEE CHESSY (2024-03-20 18:47:00)

Trajets indirects avec correspondance :
Aucun trajet indirect trouvé.
```

Figure 10: Affiche des trajets qui arrivent vers 18h00



```
Ville de départ : MONTPELLIER SAINT-ROCH
Ville d'arrivée : MARNE LA VALLEE CHESSY
Heure de départ souhaitée (format HH:MM) ou laisser vide : 15:00
Heure d'arrivée souhaitée (format HH:MM) ou laisser vide :
Trajets directs :
- MONTPELLIER SAINT-ROCH (2024-03-20 05:28:00) -> MARNE LA VALLEE CHESSY (2024-03-20 09:17:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 05:28:00) -> MARNE LA VALLEE CHESSY (2024-03-20 09:17:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 05:28:00) -> MARNE LA VALLEE CHESSY (2024-03-20 09:17:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 05:31:00) -> MARNE LA VALLEE CHESSY (2024-03-20 09:17:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 05:32:00) -> MARNE LA VALLEE CHESSY (2024-03-20 09:17:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 05:32:00) -> MARNE LA VALLEE CHESSY (2024-03-20 09:17:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 05:32:00) -> MARNE LA VALLEE CHESSY (2024-03-20 09:17:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 05:32:00) -> MARNE LA VALLEE CHESSY (2024-03-20 09:17:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 05:32:00) -> MARNE LA VALLEE CHESSY (2024-03-20 09:17:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 06:31:00) -> MARNE LA VALLEE CHESSY (2024-03-20 10:24:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 08:59:00) -> MARNE LA VALLEE CHESSY (2024-03-20 12:47:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 08:59:00) -> MARNE LA VALLEE CHESSY (2024-03-20 12:47:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 14:51:00) -> MARNE LA VALLEE CHESSY (2024-03-20 18:47:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 14:51:00) -> MARNE LA VALLEE CHESSY (2024-03-20 18:47:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 14:51:00) -> MARNE LA VALLEE CHESSY (2024-03-20 18:47:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 14:51:00) -> MARNE LA VALLEE CHESSY (2024-03-20 18:47:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 14:51:00) -> MARNE LA VALLEE CHESSY (2024-03-20 18:47:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 14:51:00) -> MARNE LA VALLEE CHESSY (2024-03-20 18:47:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 14:51:00) -> MARNE LA VALLEE CHESSY (2024-03-20 18:47:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 14:51:00) -> MARNE LA VALLEE CHESSY (2024-03-20 18:47:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 14:51:00) -> MARNE LA VALLEE CHESSY (2024-03-20 18:47:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 14:51:00) -> MARNE LA VALLEE CHESSY (2024-03-20 18:47:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 14:51:00) -> MARNE LA VALLEE CHESSY (2024-03-20 18:47:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 14:51:00) -> MARNE LA VALLEE CHESSY (2024-03-20 18:47:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 14:51:00) -> MARNE LA VALLEE CHESSY (2024-03-20 18:47:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 14:51:00) -> MARNE LA VALLEE CHESSY (2024-03-20 18:47:00)
- MONTPELLIER SAINT-ROCH (2024-03-20 14:51:00) -> MARNE LA VALLEE CHESSY (2024-03-20 18:47:00)
```

Figure 11: On affiche tous les trains qui partent avant 15h00

7. Dump de la BDD

Pour visualiser la BDD on réalise un dump avec ce que nous permet notre machine (c'est-à-dire 100,000 lignes). Nous avons essayé d'exporter nos 337,000 lignes mais alors le bouton exporter ne fonctionne plus. 100,000 maximum à l'lignes exportation.