

BASE DE DONNÉES AVANCÉES
GLO-4035

Étude de faisabilité

Rapport technique

Préparé par : Équipe 25

Catalina Chirila	536 834 846
Anne-Gabriel Dallaire	111 061 983
Thomas Gerbaud	537 188 181
Ghislain Godbout	111 250 254
Mathieu Labonté	536 994 236

Université Laval

Le 3 octobre 2023

Table des Matières

1 - Introduction	3
Explication de la problématique	3
Sommaire de la solution	3
Présentation du rapport	3
2 - Stratégie d'acquisition des données	4
Source et méthode d'extraction	4
Présentation d'exemples de données sources	4
3 – Technologies utilisés	5
Langage	5
Bases de données	5

1 - Introduction

Explication de la problématique

Sommaire de la solution

Présentation du rapport

2 - Stratégie d'acquisition des données

Source et méthode d'extraction

La ville de Montréal a été sélectionnée comme emplacement pour le prototype en raison de son excellent réseau cyclable et de ses nombreux établissements de restaurations. La ville de Montréal offre plusieurs jeux de données via son site¹ internet, notamment sur son réseau cyclable et ses établissements alimentaires. Les informations sont disponibles au grand public sous plusieurs formats téléchargeables, dont le GeoJSON qui sera utilisé pour l'élaboration du prototype. Ce format s'avère être le plus pertinent, puisqu'il permet de facilement utiliser les données géospatiales, autant pour l'emplacement des restaurants que pour les tracés du réseau cyclable. Le fichier contenant les données du réseau cyclable comprend 9 042 entrées. Pour chaque entrée, plusieurs champs qui nous seront utiles sont présents comme l'identifiant (ID_CYCL), le type de voie (TYPE_VOIE_CODE et TYPE_VOIE_DESC), sa longueur (LONGUEUR) et ses coordonnées. Le fichier contenant les établissements alimentaires comprend 47 527 entrées. Parmi elles, environ 6 000 sont des restaurants encore actifs. Pour ce jeu de données, le champ de l'identifiant (business_id), du nom (name), du type (type), du statut (statut) et de l'emplacement seront très utiles.

Présentation d'exemples de données sources

Exemple d'entrée d'un chemin de piste cyclable :

```
{ "type": "Feature", "properties": { "ID_CYCL": 26665, "ID_TRC": 0, "AFFICHEUR_DYNAMIQUE": null,
"AVANCEMENT_CODE": "E", "AVANCEMENT_DESC": "Voie cyclable existante", "COMPTEUR_CYLISTE": null,
"LONGUEUR": 23, "NBR_VOIE": 2, "NOM_ARR_VILLE_CODE": "LC", "NOM_ARR_VILLE_DESC": "Lachine", "PROTEGE_4S":
null, "REV_AVANCEMENT_CODE": null, "REV_AVANCEMENT_DESC": null, "ROUTE_VERT": null, "SAISONS4": null,
"SAS_VELO": null, "SEPARATEUR_CODE": null, "SEPARATEUR_DESC": null, "TYPE_VOIE_CODE": "5",
"TYPE_VOIE_DESC": "Piste cyclable en site propre", "TYPE_VOIE2_CODE": "1", "TYPE_VOIE2_DESC": "Non
déterminé", "VILLE_MTL": null }, "geometry": { "type": "LineString", "coordinates": [ [
-73.679351942891202, 45.428084494569902 ], [ -73.6793046063204, 45.428082228493899 ], [
-73.679277614405905, 45.428080201255497 ], [ -73.679220148686696, 45.428074712440598 ], [
-73.679142223643794, 45.428066194995203 ], [ -73.679124507891899, 45.428064332000503 ], [
-73.679101308319304, 45.428061218636898 ], [ -73.679058071957897, 45.4280553469155 ] ] ] },
```

Exemple d'entrée d'un restaurant :

```
{ "type": "Feature", "properties": { "business_id": 7828, "name": "PIZZA CITY", "address": "1821 Rue
Ontario Est", "city": "Montréal", "state": "Québec, Canada", "type": "Restaurant service rapide",
"statut": "Ouvert", "date_statut": "20090302", "latitude": "45.526212", "longitude": "-73.55913", "x":
300181.15, "y": 5042923.36 }, "geometry": { "type": "Point", "coordinates": [ -73.559129622994462,
45.526212391670889 ] } },
```

Durant la phase de développement, quelques spécificités au niveau des jeux de données seront à prendre en considération. D'abord le format GeoJSON, puisque les coordonnées géospatiales sont représentées sous la forme [longitude, latitude], ce qui est l'inverse de la convention habituelle de [latitude, longitude]. Aussi, dans le jeu de données des établissements alimentaires, un filtre devra être fait puisque certains établissements ne sont pas des restaurants et certains restaurants ne sont

¹ Site web des données ouvertes de la Ville de Montréal. (s. d.). Montréal Données Ouvertes. <https://donnees.montreal.ca/>

plus actifs. Enfin, une vérification du champ “address” devra être effectuée puisque certaines entrées (établissements fermés) ont la valeur *null* pour ce champ.

3 – Technologies utilisés

Langage

L’application sera développée en langage JavaScript, à travers l’environnement Node.js. Celui-ci est largement utilisé et a l’avantage de très bien s’accorder avec plusieurs bases de données NoSQL.

L’architecture modulaire de Node.js permet de décomposer l’application et d’ajouter simplement de nouvelles fonctionnalités, sans avoir à retravailler tout le code. JavaScript étant un langage full stack et Node.js bénéficiant d’une importante et active communauté, ainsi que de nombreuses bibliothèques, il sera plus aisé pour les intervenants futurs d’apporter des modifications selon leurs besoins. Node.js applique de surcroît une politique de support sur le long terme (LTS).

Enfin, Node.js est basé sur une architecture orientée événements, avec un système d’opérations asynchrones et non-bloquantes qui la rend capable de gérer de nombreuses requêtes simultanément. Ce choix permet d’améliorer les performances de l’API à développer et d’anticiper un trafic futur important. Finalement, Node.js assure une prise en charge multi-plateforme facilitant ainsi le déploiement de l’application sur tous les supports.

Bases de données

Le choix des bases de données s’est arrêté sur MongoDB et Neo4J. Les critères d’évaluation sont basés sur l’utilisation, le type de données, le langage de développement, ainsi que la fiabilité, la maintenabilité et l’extensibilité.

MongoDB

MongoDB est un SGBD NoSQL orienté document, largement utilisé dans l’industrie pour gérer des données semi-structurées ou non structurées. Ce choix est appuyé par les raisons principales suivantes :

Langage : MongoDB rend disponible le Driver MongoDB Javascript(Node.js).

Fiabilité : Assure une haute disponibilité grâce à des répliqués distribués et des mécanismes de reprise après des incidents autogérés, ce qui garantit une fiabilité élevée pour nos données.

Maintenabilité : Offre des schémas dynamiques, tel que le « Versioning Pattern », ce qui signifie que nous pouvons facilement ajuster la structure de nos documents en fonction de l’évolution. Cette flexibilité simplifie la maintenabilité à long terme de notre base de données.

Neo4J

Neo4j est un SGBD de graphe, idéal pour la gestion de données basées sur des relations complexes.

Requête géospatiales avancées : Propose une bibliothèque Spatiale, qui permet d’effectuer des opérations telles que la recherche de points d’intérêt à proximité et le calcul d’itinéraires optimaux.

Intégration : Permet d’intégrer des données provenant de diverses sources géospatiales, telles que des API cartographiques, des systèmes de navigation GPS et des BD géographiques externes.

Extensibilité : Prends en charge l’évolutivité horizontale, ce qui nous permet de gérer des charges de travail croissantes liées à la géolocalisation et aux itinéraires en ajoutant simplement plus de nœuds à notre cluster.