

Remise 1 : Évaluer la faisabilité

Afin de choisir la structure nécessaire à la mise en place de notre application, il est essentiel d'identifier clairement les besoins de ce système. Il importe ainsi de connaître les informations que l'utilisateur fournira à l'application afin d'exécuter les requêtes pour obtenir le résultat voulu.

Dans un premier temps, nous restreindrons l'utilisation de l'application aux pistes cyclables de la ville de Québec. L'application doit pouvoir suggérer un parcours cyclable rencontrant certains types de restaurants pour un nombre d'arrêts spécifiés et une distance maximale à parcourir. L'utilisateur doit donc fournir un lieu de départ (nom du lieu et/ou ses coordonnées), le type de restaurants recherchés (ex: cuisine française, thaï), un nombre d'arrêts ainsi que la distance maximale qu'il veut parcourir. L'application doit alors retourner une série de lieux (nom du lieu et/ou ses coordonnées) ordonnancés afin de décrire un trajet de la longueur maximale spécifiée, de même que les restaurants (nom du restaurant et/ou ses coordonnées) situés sur ce parcours pour le nombre d'arrêts souhaités.

Les bases de données supportant cette application doivent donc contenir minimalement les informations suivantes:

Les pistes cyclables	Les restaurants
Nom du lieu	Nom du restaurant
Coordonnées des points	Adresse du restaurant
	Coordonnées du restaurant
	Type de cuisine

L'application doit pouvoir calculer le parcours pour atteindre la distance maximale demandée, identifier les restaurants du type voulu situés à moins de 50 mètres du parcours et obtenir le nombre d'arrêts spécifiés.

Stratégie d'acquisition des données

Méthode d'acquisition et provenance des données

Nous utiliserons les données publiques de la ville de Québec. Cette dernière fournit les données géographiques de ses pistes cyclables : https://www.donneesquebec.ca/recherche/fr/dataset/vque_24

Pour les données concernant les restaurants, nous utiliserons une méthode de scrapping sur les pages du site <https://www.restoquebec.ca/>.

Format des données

Les données concernant les pistes cyclables de la ville de Québec sont disponibles dans plusieurs formats : CSV, GEOJSON, KML, SHP. Nous utiliserons le format GEOJSON car celui-ci est répandu, adapté aux applications web et supporté par la plupart des logiciels. Nous aurions pu choisir d'utiliser le format Shapefile qui est le plus répandu des formats de données géospatiales. Mais celui-ci implique à minima 3 fichiers (.SHP, .DBF, .SHX). De plus, nous devons générer le fichier des données restaurants à partir des données brutes récupérées, le format GEOJSON est plus facile à appréhender.

Stratégie de déploiement

Serveur web

Il y a une grande variété de choix concernant les serveurs web. Nous avons considéré les plus connus que sont Apache, le plus ancien et toujours le plus utilisé, et Nginx, plus récent et en très forte croissance. Les comparaisons montre que tout deux ont des performances équivalentes sur la plupart des aspect. En revanche Nginx est sans conteste plus performant pour le rendu se contenu statique. Nous choisirons Nginx.

Base de données

Pour les besoin de l'application nous devons stocker:

- Les données utilisateurs de bases (nom, prénom, image de profil ...)
- Les données spatials pour les pistes cyclables et les restaurants de la ville de Québec
- Les données spécifiques concernant les restaurants (type, adresse postale, description ...)

Les données utilisateurs seront stockées dans une base de données de type clé-valeur. Ce sont les données de bases qui doivent être chargées dès l'ouverture de la l'application par l'utilisateur. Dans une BD de ce type, les clés, ici l'identifiant des utilisateurs, sont stockées en mémoire vive permettant ainsi un accès optimisé aux données. De même nous stockerons dans une BD similaire les données des restaurants. Nous considérons la BD Amazone DynamoDB pour les stockage des ces données.

Pour les données spatials, la encore plusieurs possibilités s'offrent à nous. Il y a le bien connu Postgres avec son module PostGIS et quelques autres extensions pour manipuler les objets imbriqués et le format JSON. Mais notre choix se porte sur MongoDB, une BD orientée document et sans structure imposée qui supporte par défaut les requêtes géospatiales. Nous y stockerons d'une part les données des pistes cyclables ainsi que celles des points correspondant à la localisation des restaurants.

Enfin nous envisageons la possibilité de stocker certaines informations sur les restaurants, comme leurs types, dans une BD orientée colonnes pour optimiser les recherches en fonction de critères particuliers. Nous considérons la BD MariaDB pour cela.

Conteneur Docker

Des images docker existent pour chacune des bases de données que nous souhaitons utiliser, ainsi que pour le serveur Nginx:

- https://hub.docker.com/_/nginx
- https://hub.docker.com/_/mongo
- <https://hub.docker.com/r/amazon/dynamodb-local/>
- https://hub.docker.com/_/mariadb

Ressources

Bases de données géospatiales :

- <http://michaelminn.net/tutorials/gis-storage/>
- https://en.wikipedia.org/wiki/List_of_geographic_information_systems_software

- <https://blog.westerndigital.com/geospatial-data-storage-infrastructure/>
- https://en.wikipedia.org/wiki/Spatial_database
- <https://www.planetwatchers.com/choosing-the-right-database-for-geoanalytics/>

Bases de données:

- <https://www.predictiveanalyticstoday.com/top-wide-columnar-store-databases/>
- <https://www.predictiveanalyticstoday.com/top-sql-key-value-store-databases/>

Serveurs web:

- <https://www.hostingadvice.com/how-to/nginx-vs-apache/>
- <https://kinsta.com/fr/blog/nginx-vs-apache/>
- <https://iweb.com/fr/blog/introduction-aux-serveurs-de-base-de-donnees>

Jeux de données :

- <https://open.canada.ca/data/en/dataset>
- https://www.donneesquebec.ca/recherche/fr/dataset/vque_24
- <http://www.ressourcesentreprises.org/faites-votre-demande-en-ligne/faites-votre-demande-marche/>
- <https://www.restoquebec.ca/s/?restaurants=parking-available-quebec&f=23&lang=en>