

Gestion des données spatiales avec DuckDB

des bases du SQL à l'analytique
géospatiale avancée



Qiusheng Wu

Gestion des données spatiales avec DuckDB

des bases du SQL à l'analytique géospatiale
avancée

Qiusheng Wu
2026

Contents

Preface	1
Introduction	3
À qui s'adresse ce livre	4
Ce que couvre ce livre	5
Tirer le meilleur parti de ce livre	6
Conventions utilisées dans ce livre	7
Téléchargement des exemples de code	8
Tutoriels vidéo et ressources supplémentaires	8
Communauté et retours	9
Remerciements	9
À propos de l'auteur	10
Licence et droits d'auteur	10
I: Fondations de DuckDB	13
1. Premiers pas avec DuckDB	15
1.1. Introduction	15
1.2. Objectifs d'apprentissage	15
1.3. Ce qui différencie DuckDB des bases de données traditionnelles	15
1.4. Quand utiliser (et ne pas utiliser) DuckDB pour le travail spatial	17
1.5. Installation de l'interface CLI de DuckDB et exécution de votre première requête	19
1.6. Installation du client Python DuckDB	21
1.7. Installation de Visual Studio Code	24
1.8. Utilisation de l'interface utilisateur DuckDB	27
1.9. Installation de l'IDE SQL DBeaver	28
1.10. Points clés à retenir	32
1.11. Exercices	33
2. SQL essentiel pour l'analyse spatiale	36
2.1. Introduction	36
2.2. Objectifs d'apprentissage	36
2.3. Jeux de données exemples	37
2.4. Configuration de l'environnement	37
2.5. Connexion à DuckDB	39
2.6. Installation des extensions	39
2.7. Lecture de fichiers CSV depuis des URL	39
2.8. Création de tables pour de meilleures performances	40
2.9. L'instruction SQL SELECT	42
2.10. Filtrage des données avec la clause WHERE	49
2.11. Correspondance de modèles avec LIKE	51
2.12. L'opérateur IN	52
2.13. L'opérateur BETWEEN	53
2.14. Combinaison de données avec les jointures SQL	54
2.15. Plans de requête et performance	61
2.16. Aggrégation des données pour les statistiques résumées	61
2.17. Instructions conditionnelles	65
2.18. Sauvegarde des résultats	66
2.19. Fonctionnalités SQL avancées pour l'analyse spatiale	67
2.20. Commentaires SQL et documentation	72

2.21. Points clés à retenir	72
2.22. Exercices	73
3. <i>Integration de DuckDB avec Python</i>	77
3.1. Introduction	77
3.2. Objectifs d'apprentissage	77
3.3. Jeux de données d'exemple	78
3.4. Installation et configuration	79
3.5. Installation et chargement des extensions	80
3.6. Lecture de données depuis plusieurs sources	81
3.7. Intégration transparente avec les DataFrames Pandas	84
3.8. Interopérabilité avec Polars	88
3.9. Conversion des résultats et formats de sortie	88
3.10. Écriture de données sur disque	91
3.11. Stockage persistant et fichiers de base de données	91
3.12. Instructions préparées et paramètres	94
3.13. Points clés à retenir	95
3.14. Exercices	97
II: Opérations sur les données spatiales	103
4. <i>Chargement des formats de données spatiales</i>	105
4.1. Introduction	105
4.2. Objectifs d'apprentissage	106
4.3. Ensembles de données exemples	107
4.4. Installation et configuration	107
4.5. Installation et chargement des extensions	108
4.6. Téléchargement des données exemples	109
4.7. Chargement de fichiers CSV avec des coordonnées	109
4.8. Chargement de fichiers JSON	113
4.9. Interrogation directe de DataFrames Pandas	116
4.10. Chargement de fichiers Parquet pour la performance	117
4.11. Chargement de fichiers GeoJSON avec des géométries spatiales	120
4.12. Chargement de Shapefiles dans des flux de travail modernes	125
4.13. Chargement de GeoParquet pour l'analyse spatiale cloud-native	128
4.14. Stratégies de performance pour le chargement de données	130
4.15. Dépannage des problèmes courants de chargement de données	131
4.16. Points clés à retenir	131
4.17. Exercices	133
5. <i>Exportation et conversion des données spatiales</i>	137
5.1. Introduction	137
5.2. Objectifs d'apprentissage	138
5.3. Ensembles de données d'exemple	139
5.4. Installation et configuration	140
5.5. Installation et chargement des extensions	141
5.6. Chargement des données d'exemple	142
5.7. Exportation vers des DataFrames Pandas	144
5.8. Exportation vers des fichiers CSV	149
5.9. Exportation vers des fichiers JSON	152
5.10. Exportation vers des fichiers Excel	155
5.11. Exportation vers des fichiers Parquet	158

5.12. Exportation au format GeoJSON	160
5.13. Exportation au format Shapefile	164
5.14. Exportation au format GeoPackage	165
5.15. Points clés à retenir	168
5.16. Exercices	169
6. <i>Operations et fonctions géométriques</i>	175
6.1. Introduction	175
6.2. Objectifs d'apprentissage	175
6.3. Jeux de données d'exemple	176
6.4. Installation et configuration	177
6.5. Connexion à DuckDB et chargement des extensions	178
6.6. Comprendre les types de géométrie	180
6.7. Travailler avec les points	183
6.8. Travailler avec les lignes	186
6.9. Travailler avec les polygones	188
6.10. Travailler avec les collections	190
6.11. Visualisation des données spatiales de NYC	191
6.12. Traitement géométrique	194
6.13. Validité et robustesse des géométries	198
6.14. Points clés à retenir	200
6.15. Exercices	202
7. <i>Requêtes et relations spatiales</i>	206
7.1. Introduction	206
7.2. Objectifs d'apprentissage	207
7.3. Jeux de données exemples	207
7.4. Installation et Configuration	208
7.5. Connexion à DuckDB et Chargement des Extensions	209
7.6. Comprendre les Relations Spatiales	210
7.7. Tester l'Identité Géométrique	210
7.8. Relations Topologiques	213
7.9. Relations Basées sur la Distance	221
7.10. Filtrer par Seuil de Proximité	223
7.11. Requêtes du Plus Proche Voisin	224
7.12. Points Clés à Retenir	224
7.13. Exercices	226
8. <i>Jointures spatiales avancées</i>	229
8.1. Introduction	229
8.2. Objectifs d'apprentissage	230
8.3. Jeux de données d'exemple	231
8.4. Installation	231
8.5. Importation des bibliothèques et configuration	231
8.6. Connexion à DuckDB	232
8.7. Jointures d'intersection	233
8.8. Jointures par distance	238
8.9. Jointure avancée	242
8.10. Transformations de systèmes de coordonnées	244
8.11. Référence des fonctions de relations spatiales	246
8.12. Points clés à retenir	247

8.13. Exercices	249
9. Visualisation interactive des donnees	254
9.1. Introduction	254
9.2. Objectifs d'apprentissage	255
9.3. Jeux de donnees exemples	255
9.4. Installation et configuration	256
9.5. Telechargement des donnees exemples	256
9.6. Connexion a DuckDB et chargement des extensions	257
9.7. Visualisation des donnees ponctuelles	258
9.8. Visualisation des donnees lineaires	261
9.9. Visualisation des donnees polygonales	264
9.10. Points cles a retenir	268
9.11. Exercices	270
10. Travail avec les tuiles vectorielles et PMTiles	274
10.1. Introduction	274
10.2. Objectifs d'apprentissage	275
10.3. Jeux de donnees exemples	276
10.4. Installation et configuration	277
10.5. Visualiser des tuiles vectorielles directement depuis des fichiers	278
10.6. Conversion de donnees vectorielles en PMTiles	282
10.7. Visualisation de PMTiles	284
10.8. Points cles a retenir	293
10.9. Exercices	295
III: Analytique géospatiale appliquée au monde réel	301
11. Analyse de l'inventaire national des zones humides des Etats-Unis	303
11.1. Introduction	303
11.2. Objectifs d'apprentissage	304
11.3. Jeux de données utilisés dans ce chapitre	305
11.4. Comprendre la source des données	308
11.5. Accéder aux données de zones humides avec DuckDB	310
11.6. Visualiser les distributions de zones humides	313
11.7. Analyse des zones humides à l'échelle nationale	317
11.8. Points clés à retenir	330
11.9. Exercices	333
12. Analyse des empreintes de batiments mondiales	336
12.1. Introduction	336
12.2. Objectifs d'apprentissage	337
12.3. A propos de l'ensemble de donnees	338
12.4. Installation et configuration	341
12.5. Installation et chargement des extensions	341
12.6. Exploration des donnees disponibles	342
12.7. Analyse regionale avec filtrage par boite englobante	345
12.8. Comparaison multi-regions	355
12.9. Filtrage des sources de donnees et evaluation de la qualite	360
12.10. Agregation spatiale basee sur une grille	366
12.11. Agregation des donnees de batiments avec les grilles hexagonales H3	371
12.12. Points cles a retenir	382
12.13. Exercices	384

13. Analyse des donnees de taxis de NYC	388
13.1. Introduction	388
13.2. Objectifs d'apprentissage	389
13.3. A propos du jeu de donnees	390
13.4. Installation	392
13.5. Import des bibliotheques	392
13.6. Installation et chargement des extensions	392
13.7. Chargement des donnees de taxi	393
13.8. Analyse temporelle	397
13.9. Chargement des donnees de reference des zones de taxi	405
13.10. Analyse spatiale	407
13.11. Analyse des flux de trajets	413
13.12. Analyse des paiements et economique	418
13.13. Analyse du comportement des passagers	422
13.14. Analyse multi-mois	423
13.15. Visualisation	426
13.16. Conseils d'optimisation des performances	432
13.17. Points cles a retenir	433
13.18. Exercices	435
14. Developpement de tableaux de bord interactifs avec Voila et Solara	438
14.1. Introduction	438
14.2. Objectifs d'apprentissage	440
14.3. Installation de Voila et Solara	440
14.4. Introduction a Hugging Face Spaces	440
14.5. Creation d'une application Voila basique	442
14.6. Creation d'une application web avancee avec Solara	452
14.7. Points cles a retenir	460
14.8. Exercices	461

Preface

Introduction

Dans un monde de plus en plus axé sur les données, la capacité de gérer et d'analyser efficacement les informations spatiales est devenue essentielle. De l'urbanisme à la surveillance environnementale, en passant par la logistique et les services personnalisés basés sur la localisation, les données géospatiales constituent la base de nombreuses applications qui influencent notre vie quotidienne. Cependant, travailler avec des données spatiales a souvent été perçu comme un domaine spécialisé et complexe, nécessitant des outils complexes et une courbe d'apprentissage abrupte.

Voici DuckDB.

DuckDB est une base de données analytique innovante conçue pour l'efficacité et la facilité d'utilisation. En tant que base de données OLAP (traitement analytique en ligne) intégrée au processus, elle fonctionne directement au sein de votre application, éliminant le besoin de déploiements de serveurs séparés et de configurations fastidieuses. Cette nature intégrée, couplée à son architecture orientée colonnes et son moteur d'exécution vectorisé, rend DuckDB exceptionnellement rapide pour les requêtes analytiques, même avec de grands ensembles de données. Initialement populaire pour ses capacités de traitement de données à usage général, l'écosystème en expansion rapide de DuckDB (notamment ses extensions pour les données spatiales) représente un changement transformateur.

Ce livre, "**Gestion des données spatiales avec DuckDB : Des bases du SQL à l'analyse géospatiale avancée**", vise à démythifier les données géospatiales et à montrer comment DuckDB permet à tous (des analystes de données et scientifiques aux développeurs et professionnels SIG) d'exploiter ses capacités avec une simplicité et une rapidité sans précédent. Nous croyons que l'analyse spatiale robuste devrait être accessible, et non confinée à des logiciels spécialisés coûteux ou à des langages de programmation complexes. Avec DuckDB, cette accessibilité devient une réalité.

Notre parcours commence par les concepts fondamentaux des données spatiales et leur représentation, établissant une base solide en SQL pour travailler avec des points, des lignes et des polygones. Au fil de notre progression, vous découvrirez comment les fonctionnalités géospatiales natives de DuckDB (améliorées par son extension compatible PostGIS) permettent des opérations sophistiquées comme les jointures spatiales, les zones tampons et les recherches de plus proches voisins à travers d'élégantes requêtes SQL. Nous explorerons diverses applications du monde réel, démontrant comment charger, transformer, analyser et visualiser des ensembles de données spatiales, vous permettant d'extraire des informations significatives à partir d'informations géographiques.

Que vous cherchiez à intégrer l'analyse spatiale dans vos pipelines de données, à effectuer des requêtes géospatiales ad-hoc rapides ou à développer des applications interactives sensibles à la localisation, ce livre vous servira de guide complet. Nous couvrirons des sujets allant de la configuration de votre environnement DuckDB et l'importation de divers formats de fichiers spatiaux (comme les Shapefiles, GeoJSON et GeoParquet) à l'exécution de tâches analytiques complexes et à l'intégration avec des outils de visualisation.

Notre objectif n'est pas seulement de vous enseigner la syntaxe, mais de cultiver une compréhension de pourquoi ces outils et techniques sont puissants. À la fin de ce livre, vous serez compétent dans l'utilisation de DuckDB comme moteur de référence pour la gestion et l'analyse des données spatiales, ouvrant de nouvelles possibilités pour vos projets et vous permettant de prendre des décisions éclairées et spatialement conscientes.

Rejoignez-nous alors que nous plongeons dans l'intersection passionnante des capacités analytiques de DuckDB et du riche monde des données géospatiales. L'avenir de l'analyse spatiale accessible est là, et il fonctionne sur DuckDB.

À qui s'adresse ce livre

Ce livre est conçu pour quiconque est confronté aux complexités de l'analyse moderne des données spatiales. Si vous avez déjà passé des heures à attendre qu'une jointure spatiale se termine, lutté pour charger de grands ensembles de données géographiques en mémoire, ou souhaité un moyen plus simple de combiner la puissance du SQL avec les opérations spatiales, ce livre est pour vous.

Vous trouverez le plus de valeur si vous êtes

Un professionnel SIG frustré par les limitations des logiciels de bureau lors de la manipulation de grands ensembles de données. Vous connaissez QGIS ou ArcGIS, mais vous devez analyser des millions d'entités, traiter des traces GPS étendues ou intégrer l'analyse spatiale dans des flux de travail automatisés.

Un data scientist ou analyste qui rencontre fréquemment des données de localisation. Vous êtes à l'aise avec Python et pandas, mais les données spatiales semblent souvent mystérieuses. Vous souhaitez incorporer des dimensions géographiques dans vos analyses sans plonger dans des logiciels SIG complexes.

Un développeur logiciel construisant des applications qui intègrent des fonctionnalités spatiales. Vous avez besoin de requêtes spatiales rapides, souhaitez éviter une infrastructure de base de données lourde et préférez travailler avec du SQL familier plutôt qu'avec des bibliothèques spatiales spécialisées.

Un chercheur ou universitaire dans des domaines comme la géographie, les sciences environnementales ou l'urbanisme. Votre recherche implique de grands ensembles de données spatiales, et vous avez besoin de méthodes d'analyse reproductibles et évolutives qui peuvent s'adapter à des volumes de données croissants.

Un professionnel de la Business Intelligence traitant des données commerciales basées sur la localisation. Qu'il s'agisse d'emplacements de magasins, d'itinéraires de livraison, de territoires clients ou de portefeuilles immobiliers, vous devez fusionner les métriques commerciales avec les informations spatiales.

Prérequis essentiels

Vous devriez être à l'aise avec :

- **La programmation Python** : comprendre les variables, les fonctions et comment importer des bibliothèques (l'expertise n'est pas requise)
- **Les concepts d'analyse de données** : filtrer les enregistrements, agréger les données et joindre les tables
- **Les fondamentaux du SQL** : familiarité avec les clauses SELECT, WHERE et GROUP BY (nous couvrirons les aspects spatiaux)
- **Les bases des données spatiales** : comprendre que les données ont une localisation (latitude/longitude, projections)

Connaissances utiles (mais non requises)

- Expérience avec pandas, GeoPandas ou les notebooks Jupyter
- Exposition préalable aux bases de données ou entrepôts de données
- Familiarité avec les logiciels SIG (QGIS, ArcGIS, PostGIS)
- Connaissance des formats de fichiers spatiaux (GeoJSON, Shapefiles, Parquet)

Si vous êtes nouveau en programmation Python

Si vous êtes nouveau dans la programmation Python géospatiale, le livre suivant fournit une excellente introduction aux concepts fondamentaux des SIG et à la programmation Python :

Wu, Q. (2025). *Introduction to GIS Programming: A Practical Python Guide to Open Source Geospatial Tools.* Independently published. ISBN 979-8286979455. <https://www.amazon.com/dp/B0FFW34LL3>

Ce que couvre ce livre

Ce livre offre un parcours structuré des bases du SQL à l'analyse géospatiale avancée, vous équipant de compétences pratiques à travers des exemples du monde réel. Chaque chapitre progresse des requêtes simples aux analyses spatiales complexes, développant votre expertise en gestion moderne des données géospatiales.

Partie I : Fondements de DuckDB (*Chapitres 1-3*)

Maîtrisez les concepts essentiels qui sous-tendent tout le contenu suivant :

- **Chapitre 1 : Premiers pas avec DuckDB** : Installation, premières requêtes et aperçu de la façon dont DuckDB révolutionne l'analyse spatiale.
- **Chapitre 2 : SQL essentiel pour l'analyse spatiale** : Modèles SQL clés pour le filtrage, l'agrégation, les jointures et l'optimisation des requêtes adaptées aux données spatiales.
- **Chapitre 3 : Intégration Python de DuckDB** : Combinez la puissance du SQL avec la flexibilité de pandas pour créer un flux de travail d'analyse spatiale fluide.

À la fin de la Partie I, vous interrogerez avec confiance des ensembles de données spatiales et intégrerez DuckDB dans tout pipeline d'analyse basé sur Python.

Partie II : Opérations sur les données spatiales (*Chapitres 4-10*)

Plongez dans la boîte à outils spatiale principale, couvrant tout, du chargement des données à l'analyse avancée :

- **Chapitre 4 : Chargement des formats de données spatiales** : Importez divers formats incluant les coordonnées CSV, GeoJSON depuis des API, des Shapefiles massifs et GeoParquet hébergé dans le cloud.
- **Chapitre 5 : Exportation et conversion de données spatiales** : Transformez vos résultats dans n'importe quel format requis par vos parties prenantes.
- **Chapitre 6 : Opérations et fonctions géométriques** : Créez, mesurez et transformez des entités spatiales en utilisant des fonctions SQL.
- **Chapitre 7 : Requêtes et relations spatiales** : Maîtrisez les équivalents spatiaux des jointures de tables : containment, intersection et proximité.
- **Chapitre 8 : Jointures spatiales avancées** : Combinez des ensembles de données par localisation plutôt que par ID, ce qui est l'essence de l'analyse spatiale.
- **Chapitre 9 : Visualisation interactive des données** : Générez des cartes et graphiques convaincants qui communiquent efficacement les récits spatiaux de vos données.
- **Chapitre 10 : Travailler avec les tuiles vectorielles et PMTiles** : Déployez des cartes interactives capables de gérer des millions d'entités en douceur.

À la fin de la Partie II, vous serez capable de gérer n'importe quel format de données spatiales, d'exécuter des opérations complexes et de créer des visualisations professionnelles.

Partie III : Analyse géospatiale du monde réel (*Chapitres 11-14*)

Explorez quatre études de cas complètes utilisant des ensembles de données réels à grande échelle :

- **Chapitre 11 : Analyse de l'inventaire national des zones humides des États-Unis** : Effectuez des analyses environnementales dans les 50 États, en traitant des millions de polygones de zones humides.
- **Chapitre 12 : Analyse des empreintes de bâtiments mondiales** : Analysez les données urbaines en utilisant les empreintes de bâtiments mondiales d'Overture Maps.
- **Chapitre 13 : Analyse des données de trajets de taxi de NYC** : Découvrez des modèles spatio-temporels à partir de centaines de millions de trajets en taxi, révélant des informations sur la mobilité urbaine.
- **Chapitre 14 : Développement de tableaux de bord interactifs avec Voilà et Solara** : Construisez et déployez des applications web qui rendent vos analyses accessibles aux parties prenantes.

À la fin de la Partie III, vous aurez des projets dignes d'un portfolio démontrant vos capacités avancées d'analyse spatiale.

Thèmes transversaux tout au long du livre

- **Performance à l'échelle** : Techniques efficaces que vous manipulez des milliers ou des millions d'entités spatiales.
- **Flux de travail cloud-natifs** : Traitez les données directement depuis S3 et intégrez-vous de manière transparente aux piles de données modernes.
- **Analyse reproductible** : Code et méthodes partageables pouvant être versionnés et déployés en production.
- **Défis des données du monde réel** : Abordez les problèmes de projection, les valeurs manquantes et les préoccupations de qualité des données.
- **Modèles d'intégration** : Combinez DuckDB avec l'écosystème Python géospatial plus large pour une fonctionnalité améliorée.

Ce qui rend ce livre différent

Contrairement aux discussions théoriques ou aux tutoriels spécifiques à un outil, ce livre met l'accent sur la *réolution de vrais problèmes*. Chaque technique est ancrée dans des défis analytiques réels, démontrée avec des ensembles de données réels et expliquée clairement en termes de quand et pourquoi l'utiliser.

Tirer le meilleur parti de ce livre

Pour maximiser votre expérience d'apprentissage avec ce livre, considérez les recommandations suivantes :

Configurez un environnement de développement approprié : Installez Python et les bibliothèques requises comme décrit au Chapitre 1. Un environnement bien configuré vous fera gagner du temps et évitera des frustrations tout au long de votre parcours d'apprentissage. Envisagez d'utiliser conda ou uv pour gérer vos paquets Python, car cela simplifie l'installation des bibliothèques géospatiales.

Suivez les exemples de code : Ce livre est conçu pour être interactif. Ne vous contentez pas de lire le code ; tapez-le, exéutez-le et expérimenez avec des modifications. La compréhension vient par la pratique, et chaque exemple développe des compétences dont vous aurez besoin plus tard.

Travaillez les exercices : Chaque chapitre inclut des exercices conçus pour renforcer les concepts que vous avez appris. Ce ne sont pas des extras optionnels ; ils font partie intégrante du processus d'apprentissage. Commencez par les exercices guidés, puis défiez-vous avec vos propres projets.

Utilisez des données réelles : Bien que le livre fournisse des ensembles de données pour les exemples et exercices, essayez d'appliquer les techniques à des données de votre propre domaine ou de vos intérêts. Cela vous aidera à comprendre comment les concepts s'appliquent à des scénarios du monde réel et renforcera votre confiance en vos capacités.

Construisez des projets : Au fur et à mesure que vous progressez dans le livre, envisagez de travailler sur un projet personnel qui vous intéresse. Il peut s'agir d'analyser des données de votre recherche, de créer des cartes pour votre communauté ou de résoudre un problème que vous avez rencontré dans votre travail.

Soyez patient avec vous-même : La programmation peut être frustrante, surtout lorsque vous apprenez. Attendez-vous à rencontrer des erreurs, à passer du temps à déboguer et à vous sentir parfois bloqué. C'est normal et fait partie du processus d'apprentissage. Faites des pauses si nécessaire, et rappelez-vous que l'expertise se développe progressivement grâce à une pratique constante. Si vous êtes bloqué, n'hésitez pas à demander de l'aide sur le dépôt GitHub du livre.

Continuez à pratiquer : Les compétences de ce livre nécessitent une pratique régulière pour être maintenues et développées. Réservez du temps régulièrement pour travailler sur des projets de programmation géospatiale, même s'ils sont petits.

Conventions utilisées dans ce livre

Ce livre utilise plusieurs conventions pour vous aider à naviguer dans le contenu et comprendre les exemples de code :

Formatage du code : Tout le code Python apparaît en police à espacement fixe dans des blocs de code. Lorsque le code apparaît dans le texte régulier, il est formaté comme ceci. Les noms de fichiers et de répertoires sont également formatés en police à espacement fixe.

Exemples de code : La plupart des exemples de code sont complets et exécutables. Ils incluent des commentaires expliquant les concepts et techniques clés démontrés. Des numéros de ligne peuvent être inclus pour référence dans le texte d'accompagnement.

```
# Ceci est un exemple de bloc de code
import leafmap
m = leafmap.Map()
m.add_basemap("OpenTopoMap") # ajouter un fond de carte à la carte
m
```

Guide de style SQL : Pour la cohérence et la lisibilité, les exemples SQL suivent ces modèles :

- **Mots-clés en MAJUSCULES** : SELECT, FROM, WHERE, JOIN
- **Noms de fonctions conservent leur casse** : ST_Area(), read_csv_auto()
- **Noms de tables et colonnes en minuscules** : cities, population
- **Indentation pour la lisibilité** : Les requêtes multi-lignes sont formatées pour plus de clarté

```
SELECT name, ST_Area(geometry) as area
FROM neighborhoods
WHERE borough = 'Manhattan'
ORDER BY area DESC;
```

Instructions en ligne de commande : Les commandes à entrer en ligne de commande ou dans le terminal sont affichées avec une invite `$` (ne tapez pas le symbole `$` lui-même) :

```
$ pip install leafmap
$ python script.py
```

Téléchargement des exemples de code

Tous les exemples de code, ensembles de données et matériels supplémentaires pour ce livre sont disponibles gratuitement sur GitHub :

<https://github.com/giswqs/duckdb-spatial>

Pour télécharger les matériels, vous pouvez utiliser l'une des méthodes suivantes :

- **Cloner le dépôt** (si vous avez Git installé) :

```
$ git clone https://github.com/giswqs/duckdb-spatial.git
```

- **Télécharger en ZIP** (si vous préférez ne pas utiliser Git) :

- Visitez la page du dépôt GitHub
- Cliquez sur le bouton vert **Code**
- Sélectionnez **Download ZIP**
- Extrayez les fichiers à l'emplacement de votre choix

- **Parcourir les fichiers individuels** en ligne via l'interface GitHub si vous n'avez besoin que d'exemples spécifiques

Le dépôt est régulièrement mis à jour avec des corrections, des améliorations et des exemples supplémentaires. Revenez périodiquement pour les mises à jour, ou **surveillez** le dépôt sur GitHub pour être notifié des changements.

Si vous trouvez des erreurs dans le code ou avez des suggestions d'amélioration, veuillez ouvrir un ticket ou soumettre une pull request sur GitHub. Les contributions de la communauté aident à améliorer cette ressource pour tout le monde.

Tutoriels vidéo et ressources supplémentaires

En complément du contenu écrit, ce livre est accompagné d'une série complète de tutoriels vidéo qui parcourent les concepts clés et fournissent des exemples supplémentaires :

<https://tinyurl.com/duckdb-spatial-videos>

Les vidéos sont conçues pour compléter, et non remplacer, le matériel écrit. Elles sont particulièrement utiles pour :

- Les apprenants visuels qui bénéficient de voir le code écrit et exécuté
- Comprendre des concepts complexes à travers des explications multiples
- Apprendre le flux de travail de développement et les meilleures pratiques
- Voir comment aborder les problèmes et déboguer les erreurs

La playlist est organisée pour suivre la structure du livre. Vous pouvez les regarder dans l'ordre au fur et à mesure que vous progressez dans le livre, ou passer à des sujets spécifiques selon vos besoins.

Les vidéos ont été créées à l'automne 2023 lorsque j'enseignais le cours **Spatial Data Management**¹ à l'Université du Tennessee. Bien que le cours soit terminé, les vidéos restent pertinentes et peuvent être utilisées comme références pour le livre. Des vidéos supplémentaires seront ajoutées à l'avenir.

Communauté et retours

Je suis ouvert aux retours, questions et suggestions des lecteurs. Vos commentaires aident à améliorer le livre et le rendent plus utile pour la communauté de programmation géospatiale.

Pour les questions et discussions liées au livre :

- GitHub Issues : <https://github.com/giswqs/duckdb-spatial/issues>
- GitHub Discussions : <https://github.com/giswqs/duckdb-spatial/discussions>

Types de retours particulièrement utiles :

- Erreurs ou explications peu claires dans le texte ou le code
- Suggestions d'exemples ou de cas d'utilisation supplémentaires
- Idées de nouveaux sujets ou chapitres
- Signalements de problèmes de compatibilité avec différents systèmes d'exploitation ou versions de bibliothèques
- Témoignages de réussite sur la façon dont vous avez appliqué les techniques du livre

Remerciements

Ce livre existe grâce aux contributions de nombreuses personnes et de la communauté géospatiale open source plus large.

Tout d'abord, je tiens à remercier l'**équipe de développement de DuckDB** pour avoir créé un système de base de données aussi exceptionnel. Leur engagement envers la performance, la simplicité et les principes open source a rendu l'analyse spatiale accessible à un public beaucoup plus large. Une reconnaissance spéciale va aux membres de l'équipe qui ont contribué aux capacités spatiales de DuckDB.

Je suis reconnaissant envers mes **collègues et étudiants** de l'Université du Tennessee qui ont fourni des retours, testé des exemples et aidé à affiner le contenu à travers le cours **Spatial Data Management**. Leurs questions et perspectives ont rendu ce livre beaucoup plus solide.

La **communauté géospatiale open source** mérite une reconnaissance particulière. Des projets comme GDAL, GeoPandas, Shapely et d'innombrables autres forment la base qui rend possible l'analyse spatiale moderne. L'esprit collaboratif de cette communauté continue d'inspirer mon travail.

Merci aux **premiers lecteurs** qui ont fourni des retours sur les chapitres préliminaires et aidé à identifier les domaines nécessitant des clarifications ou des améliorations. Leurs perspectives diverses (des profes-

¹<https://geog-414.gishub.org>

sionnels SIG chevronnés aux nouveaux venus en science des données) ont aidé à garantir que ce livre serve son public cible.

Je tiens à remercier **ma famille** pour sa patience et son soutien pendant les nombreuses soirées et week-ends passés à écrire. Leur compréhension et leurs encouragements ont rendu ce projet possible.

Enfin, merci à **vous**, le lecteur, pour votre intérêt pour l'analyse des données spatiales et les outils open source. Ce sont des praticiens comme vous qui stimulent l'innovation et font du domaine géospatial un endroit si passionnant pour travailler.

Si ce livre vous aide à résoudre des problèmes spatiaux plus efficacement, alors tous les efforts en valaient la peine.

À propos de l'auteur

Dr. Qiusheng Wu est professeur associé au Département de Géographie et Développement Durable de l'Université du Tennessee, Knoxville. Il est également Amazon Scholar. Les recherches du Dr Wu se concentrent sur l'avancement de l'analyse géospatiale open source à travers le cloud computing et la GeoAI. Il est le créateur et mainteneur de plusieurs paquets Python open source largement utilisés, notamment [Geemap](#)², [Leafmap](#)³, [SAMGeo](#)⁴ et [GeoAI](#)⁵, qui intègrent des plateformes géospatiales cloud avec une analyse et une visualisation alimentées par l'IA. Le travail du Dr Wu fait le pont entre la télédétection, l'observation de la Terre et l'intelligence artificielle pour rendre les données géospatiales à grande échelle plus accessibles, reproductibles et intelligentes pour les chercheurs, les éducateurs et les praticiens du monde entier. Ses projets open source sont disponibles sur GitHub à <https://github.com/opengeos>.

Licence et droits d'auteur

Ce livre adopte les principes de la science ouverte et de l'éducation ouverte. Pour soutenir la transparence, l'apprentissage et la réutilisation, les **exemples de code** de ce livre sont publiés sous licence [Creative Commons Attribution 4.0 International \(CC BY 4.0\)](#). Cela signifie que vous êtes libre de copier, modifier et distribuer le code, même à des fins commerciales, tant qu'un crédit approprié est donné.

Veuillez attribuer l'utilisation du code en citant le livre ou en créant un lien vers le dépôt GitHub :

Wu, Q. (2025). *Spatial Data Management with DuckDB: From SQL Basics to Advanced Geospatial Analytics*. Independently published. PDF edition ISBN 979-8993859705; Print edition ISBN 979-8274710572. <https://duckdb.gishub.org>

Bien que le code soit librement disponible, le **texte, les figures et les images** de ce livre sont **protégés par le droit d'auteur** par l'auteur et ne peuvent être reproduits, redistribués ou modifiés sans autorisation explicite. Cela inclut tout le contenu écrit, les diagrammes personnalisés et les visualisations intégrées, sauf indication contraire.

Si vous souhaitez réutiliser ou adapter tout matériel non-code du livre (par exemple, pour l'enseignement, des présentations ou des publications), veuillez contacter l'auteur pour demander une autorisation.

²<https://geemap.org>

³<https://leafmap.org>

⁴<https://samgeo.gishub.org>

⁵<https://opengeoai.org>

Cette approche de double licence aide à équilibrer l'accès ouvert aux matériels d'apprentissage avec la protection du travail créatif original. Merci de respecter ces conditions et de soutenir la communauté géospatiale open source.