Engineer

GDAL

бібліотека для читання й перетворення геоданих

Що це таке

GDAL (Geospatial Data Abstraction Library; часто пишуть GDAL/OGR) — це програмне забезпечення з відкритим кодом, яке відкриває геодані, збережені в одному форматі, і записує їх в іншому, а за потреби ще й перепроєктовує дорогою. Власна документація проєкту називає його «translator library», тобто бібліотекою-перекладачем.

Геодані бувають двох видів. Растрові — це сітка комірок: супутниковий знімок, аерофото, цифрова модель рельєфу. Векторні — точки, лінії й полігони з атрибутами: адреси, дороги, межі земельних ділянок. GDAL дає програмі, яка його викликає, одну абстрактну модель для кожного виду, а під моделлю стоїть окремий драйвер для кожного формату. GeoTIFF, JPEG 2000 і NetCDF належать до растрової частини, Shapefile, GeoJSON, GeoPackage, KML і таблиці PostgreSQL — до векторної, і форматів обох видів значно більше. Програма, написана поверх GDAL, читає їх усі тими самими викликами.

Векторна частина багато років називалася OGR, і дві половини мали окремі структури драйверів, доки GDAL 2.0 їх не об'єднав. Тому векторні утиліти досі носять старий префікс: `ogrinfo` описує векторний набір даних, а `ogr2ogr` його конвертує. Роботу над бібліотекою почав Френк Вармердам наприкінці 1998 року. Вона поширюється під ліцензією в стилі MIT і є проєктом Open Source Geospatial Foundation (OSGeo).

Користуються нею трьома способами. З командного рядка: `gdalinfo`, `gdal_translate` і `gdalwarp` для растрів, `ogrinfo` і `ogr2ogr` для векторів. З версії 3.11 є ще й єдина команда `gdal` з підкомандами на кшталт `gdal raster` і `gdal vector`, яку проєкт досі позначає як попередню. З коду — через прив'язки до Python та інших мов. І всередині іншого програмного забезпечення: QGIS, наприклад, читає векторні формати через драйвери

GDAL і виставляє утиліти GDAL у своєму наборі інструментів Processing. Типова конвертація вміщується в один рядок; тут shapefile стає GeoPackage і водночас перепроєктовується в зону UTM 36N, у якій лежить Київ:

```
ogr2ogr -f GPKG -t_srs EPSG:32636 dilianky.gpkg dilianky.shp
```

На момент написання поточний випуск — 3.13.3. Конкурентів на цьому рівні небагато. ArcGIS має власні інструменти конвертації, а для окремих форматів існують невеликі самостійні читачі, але команда, якій треба приймати те, що надсилає замовник, зазвичай так чи інакше отримує GDAL у своєму стеку, часто тому, що на ньому побудована вже обрана бібліотека.

Застосуйте поняття до рішення

Застосуйте «Що таке GDAL?» до конкретного рішення. Ми можемо пов'язати поняття із системою, яку ви створюєте або купуєте. Усе, що відбувається в певному місці, — судна, транспорт, ділянки, датчики — на карті, де можна шукати, вимірювати й ділитися.

Переглянути наші послуги Зв'язатися з нами →

Перевірте припущення за терміном

Використайте пояснення та пов'язані роботи, щоб сформулювати запитання для проєкту. Ми можемо перевірити припущення разом до вибору інструмента. Прискорила пайплайн географічних даних у 50 разів — 50x швидші пайплайни географічних даних

Прискорила пайплайн географічних даних у 50 разів Зв'язатися з нами →

Чому це важливо

Кожна система геоданих має межу, де дані ззовні потрапляють усередину: shapefile від міської ради, GeoTIFF від геодезиста, експорт GeoJSON від партнера, файл KML, який хтось накреслив у браузері. Рішення тут — хто стоїть на цій межі. Це може бути людина, яка проклацує настільну програму, або скрипт, що щоразу робить те саме. GDAL робить такий скрипт дешевим.

Конвертацію, записану як команда `ogr2ogr` чи `gdalwarp`, можна прочитати під час рев'ю, тримати в системі контролю версій і запустити знову, коли прийде наступна поставка. Конвертацію вручну роблять один раз, і робить її той, хто пам'ятає кроки. Для даних, що надходять частинами місяць за місяцем, саме ця різниця визначає, чи втримається якість.

Ліцензія теж має значення. Платити не треба ні за користувача, ні за машину, погодження постачальника не потрібне, і той самий інструмент працює на ноутбуці, на сервері збірки й у контейнері.

Третя причина — доступ до сховища. Віртуальні файлові системи GDAL дозволяють драйверу читати дані там, де вони лежать. З префіксом `/vsicurl/` файл читається через HTTP запитами окремих діапазонів байтів, тож більшість драйверів може читати віддалений набір даних частинами, не завантажуючи його повністю, а `/vsis3/` робить те саме з Amazon S3. У парі з Cloud Optimized GeoTIFF — файлом, розбитим на плитки, з власними оглядовими рівнями й компонуванням, що тримає кількість запитів низькою, — растр на багато гігабайтів може лежати в об'єктному сховищі й читатися по одному фрагменту.

Витрати теж реальні. GDAL — велика нативна бібліотека з багатьма необов'язковими залежностями, і набір драйверів у конкретній інсталяції залежить від того, як її зібрано. Дві машини, на кожній з яких «є GDAL», можуть відкривати різні формати. А ще він перекладає сумлінно, і це означає, що так само сумлінно він переносить усі обмеження формату, у який його просять писати.

Як це виглядає на практиці

GDAL трапляється у двох проєктах на цьому сайті, і в кожному в іншій ролі.

З 2015 до 2022 року роль керівниці команди розробки GIS-програмного забезпечення охоплювала понад тридцять GIS-проєктів від розробки й розгортання до підтримки. В основі лежав PostgreSQL із PostGIS для просторових даних, а GDAL/OGR переносив їх між форматами: shapefile, GeoJSON, GeoTIFF, KML і векторні плитки. QGIS і JOSM служили для самої роботи з даними, а вебкарти над ними будувалися на Mapbox GL і Leaflet. Цим переліком форматів обслуговували весь діапазон робіт: цифрову картографію у 2D і 3D, обробку LiDAR, георектифікацію, векторизацію, карти приміщень і навігацію в них. Кожен проєкт після запуску підтримувався в продакшні.

З 2022 до 2024 року, на посаді інженерки даних у масштабному проєкті геопросторової аналітики, GDAL був одним з інструментів пошуку й виправлення помилок. Просторові помилки діагностувалися й виправлялися за допомогою GDAL, QGIS та ArcGIS, а оскільки ті самі завдання з очищення поверталися знову й знову, їх автоматизували скриптами на Bash. PostGIS відповідав за просторові запити й індексацію, процедури на PL/pgSQL і Transact-SQL перетворювали дані в PostgreSQL і SQL Server, а

очищені дані потрапляли в інтерактивні візуалізації на Mapbox. Було усунено понад 1 000 проблем, і ця робота сприяла скороченню часу виконання просторових запитів на 35%.

Де виникають проблеми

GDAL робить саме те, що йому сказали. Більшість проблем з'являється тоді, коли сказали трохи не те, що мали на увазі.

Точки опиняються не на тому континенті. Починаючи з GDAL 3.0 бібліотека дотримується порядку осей, який визначив орган, що веде систему координат, і для EPSG:4326, звичної системи широти й довготи, першою йде широта. Код, написаний під GDAL 2, передає першою довготу, і GeoJSON теж. Якщо змішати ці два підходи, кожна пара міняється місцями, і точка в Києві з'являється на південному заході Ірану. Оберіть один порядок на межі програми й задокументуйте його. У коді, якому потрібна довгота першою, задокументований перемикач — `SetAxisMappingStrategy(OAMS_TRADITIONAL_GIS_ORDER)` на об'єкті просторової прив'язки.

Назви колонок обрізаються, а дати втрачають час. Shapefile обмежує назви атрибутів 10 символами. GDAL скорочує довші назви під час запису, а коли дві скорочені назви збігаються, урізає їх до 8 символів і додає порядковий номер, тож `abcdefghijkl` перетворюється на `abcdefgh_1`. Формат зберігає дати без часу. Для сумісності з іншим програмним забезпеченням GDAL рекомендує, щоб розмір файлів `.shp` і `.dbf` не перевищував 2 ГБ. Кирилиця може стати нечитабельною, якщо кодування визначено неправильно за відсутності файла `.cpg`, що вказує кодування. Пишіть GeoPackage, де одержувач це дозволяє. Де shapefile обов'язковий, перейменуйте колонки свідомо ще до конвертації.

Віддалений растр завантажується цілком. Якщо дати GDAL звичайну адресу `https://` без префікса, він обере свій HTTP-псевдодрайвер, який завантажує весь файл у пам'ять, перш ніж його прочитає справжній драйвер. Для маленького файлу це годиться, для ортофотоплану всієї країни — ні. Додайте перед адресою `/vsicurl/`, щоб драйвер брав лише потрібні діапазони байтів, і зберігайте великі растри як Cloud Optimized GeoTIFF, щоб цих діапазонів було мало. А потім не чіпайте файл: редагування COG на місці ламає його компонування, і з GDAL 3.8 спроба відкрити його для оновлення відхиляється, якщо явно не дозволити таке порушення.

Числа з CSV приходять як текст. Драйвер CSV читає кожне поле як текст, якщо не сказати інакше, тож колонка з чисельністю населення сортується за абеткою, а сума десь далі в ланцюжку завершується помилкою типу. Увімкніть параметр відкриття `AUTODETECT_TYPE=YES` або покладіть поруч із CSV файл `.csvt` з типом кожної колонки.

Оновлення ламає імпорт у Python. Пакет `gdal` на PyPI збирається під уже встановлений на машині GDAL, тож версії мають збігатися. Інструкція проєкту зі встановлення прив'язує одну до одної:

```
pip install gdal[numpy]=="$(gdal-config --version).*"
```

Що почитати далі

- GDAL documentation — власний опис бібліотеки-перекладача, її драйверів і ліцензії.
- Programs — усі утиліти командного рядка, від `gdalinfo` і `ogr2ogr` до нової команди `gdal`.
- GDAL Virtual File Systems — `/vsicurl/`, `/vsis3/` та інші префікси для читання даних там, де вони лежать.
- OGR Coordinate Reference Systems and Coordinate Transformation tutorial — з розділом про порядок осей і те, що змінилося в GDAL 3.0.
- COG – Cloud Optimized GeoTIFF generator — компонування файлу, оглядові рівні й чому редагування на місці їх ламає.

Словник

Обирайте за фактичними умовами

Опишіть робочий процес, наявні інтерфейси та користувачів, яким потрібна зміна. Ми можемо визначити перевірюваний результат і спосіб передачі, який дасть вашій команді змогу підтримувати його.

Дивитися, скільки це коштує Зв'язатися з нами →

Зробіть термінологію практичною

Запитайте, як «Що таке GDAL?» стосується вашої організації. Конкретного прикладу з поточної системи достатньо для початку розмови.

Переглянути наші послуги

welcome@engineer.company +45 50 10 61 84 Копенгаген, Данія

Візьміть цю сторінку із собою

Візитівка

Дбайливо зроблено з любов'ю — D.

CVR 45969703 · Tagensvej 105C, 3. th, 2200 København N

Авторське право © 2025 – дотепер · Engineer ApS · Усі права захищено.

Надруковано з https://engineer.company/uk/glossary/gdal/