



Открытые
ГИС 2015

GISconf 2015. Николай Лебедев

Обзор актуальных форматов векторных геоданных

Формат

Определённая структура
информационного объекта

Спецификация структуры данных,
записанных в компьютерном файле

Геоданные

Данные, содержащие пространственную информацию

- Растровые (содержат значения ячеек сетки)
- Векторные (содержат координаты объектов)

№1 Shapefile

- Представлен в ArcView 2 в 1993г.
- Координаты отдельно от атрибутов
- Тип координат double (8 байт на каждую)
- Shapefile это не файл, а группа файлов!

3 обязательных файла:

.shp - бинарная геометрия объектов

.dbf - табличные данные атрибутов, формат dBase III

.shx - файл связи индекс объектов

Shapefile и его свита

Необязательные файлы в составе Shapefile:

- .cpg - кодировка таблицы .dbf
- .sbn и .sbx - пространственные индексы объектов
- .fbn и .fbx - пространственные индексы read-only объектов
- .ain и .aih - индексные файлы атрибутивных таблиц
- .ixs - индекс геокодирования
- .mxs - индекс геокодирования (ODB format)
- .atx - атрибутивный индекс для каждой колонки таблицы для ArcGIS >= 8
- .qix - пространственный индекс типа quadtree, используемый GDAL/OGR и MapServer
- .shp.xml - метаданные по схеме ISO 19115 или др. XML схемы
- .prj – файл с описанием системы координат

Маленькие слабости

- dBase III (1986)
- Необходима поддержка синхронизации .shp и .dbf
- Кодировка таблицы атрибутов в .cprg
- Система координат передается в отдельном .prj файле
- Не более 10 символов в названии поля (featurecla)
- Не более 255 символов в строке
- Не более 2GB .shp

Почему все в шейпах?

Потому что шейп открытый

Потому что шейп бинарный и с точностью 15 знаков

Потому что шейп быстрый (dBase, .shx)

Потому что шейп универсальный

Point, Polyline, Polygon, MultiPoint, PolylineZ, PolygonZ, MultiPointZ, PointM, PolylineM, PolygonM, MultiPointM, MultiPatch

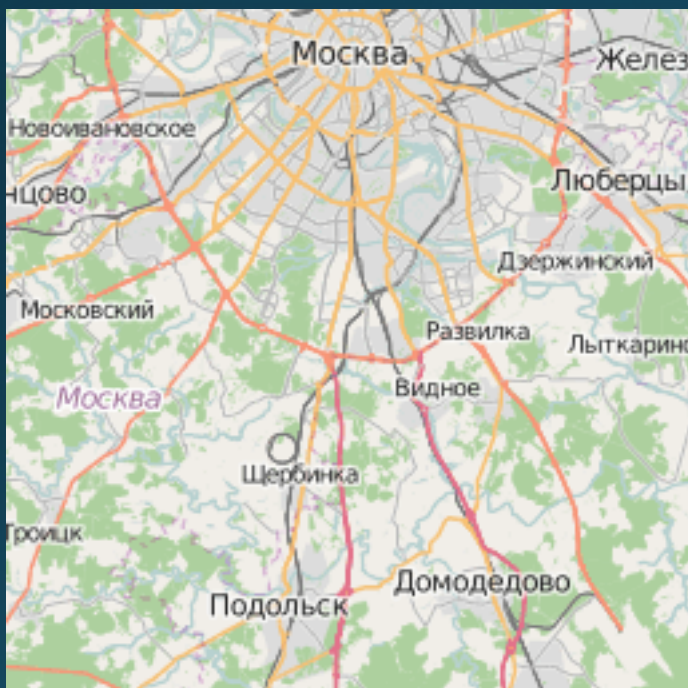
Потому что он здесь давно

А как же web?

shapefile-js - JavaScript библиотека для чтения zip архива с шейпфайлом

Web-карты работают на растровых тайлах (?)

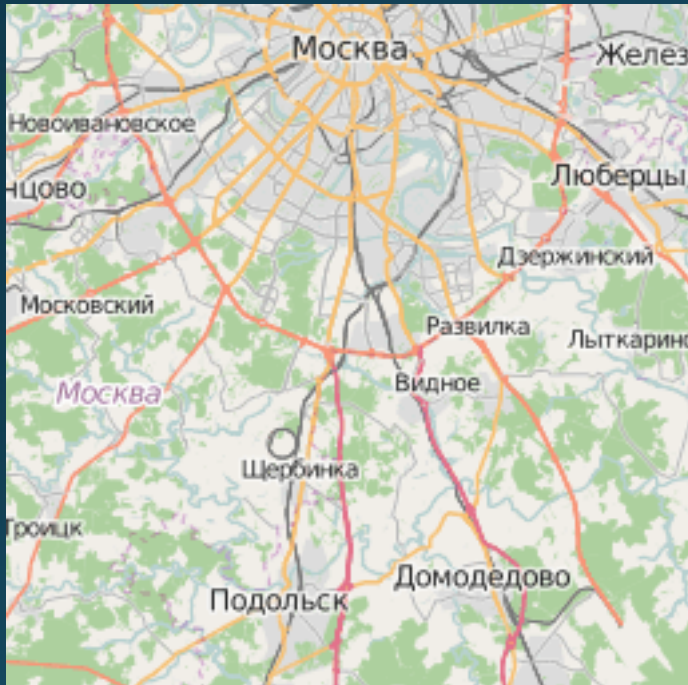
Сколько весит один тайл OSM



~

?

Сколько весит один тайл OSM



~ 40 kb

Растровый web поднимает вопросы

- Тайлов довольно много
 $6 * 4 * 40kb = 960kb$ (на один экран)
- Как обеспечить интерактивность?
 - UTFGrid - текстовая матрица кодированных значений
 - WFS - OGC-сервис для получения пространственных объектов
- Как наложить свои данные на карту OSM и не закрыть при этом надписи на карте?

Векторные бонусы

- WebGL стилизация (не надо кэшировать разные стили одних и тех же данных)
- Управление порядком отображения данных
- Overzoom (не надо запрашивать на сервере данные для более крупных масштабов, если детализация уже максимальная)
- Интерактивность
- Пространственные операции в браузере (Turf.js, D3.js)

Как передать векторные данные в браузер?

<h1>Привет GML</h1>

- Geography Markup Language (GML)

Язык разметки для описания пространственных объектов (XML схема)

Только 2D геометрия

- Keyhole Markup Language (KML)

Расширение GML от Google

Добавляет описание стилизации

Добавляет описание 3D моделей

Мы все - объекты javascript

GeoJson

- Структурированный текстовый формат JSON
- Несколько типов геометрии в одном файле
- Нефиксированный набор свойств

```
{ "type": "Feature",  
  "geometry": {  
    "type": "LineString",  
    "coordinates": [[30, 10], [10, 30], [40, 40]]  
  }  
  "properties" : {...}  
}
```

Вперед в прошлое

ARC/Info Eoo (80e годы)

Point, Arc, Route, Polygon
(почти как node, way и relation в OSM)



Вперед в прошлое

ARC/Info Eoo (80e годы)

Point, Arc, Route, Polygon

(почти как node, way и relation в OSM)

ТопоJSON - сохраняется

топология

Поддерживается GDAL ≥ 1.11

Quantization

Simplification



ISO/IEC 13249-3:2011

WKT – Well Known Text

- Разработан OGC, утвержден как стандарт ISO (ISO/IEC 13249-3:2011 standard, "Information technology -- Database languages -- SQL multimedia and application packages -- Part 3: Spatial" (SQL/MM))
- Point, LineString, Polygon, MultiPoint, MultiLineString, MultiPolygon, GeometryCollection, CircularString, CompoundCurve, CurvePolygon, MultiCurve, MultiSurface, Curve, Surface, PolyhedralSurface, TIN, Triangle
- 2D (x, y), 3D (x, y, z), 4D (x, y, z, m) геометрия

POINT (30 10)

LINESTRING (30 10, 10 30, 40 40)

POLYGON ((30 10, 40 40, 20 40, 10 20, 30 10))

Бинарный мир

- WKB – Well Known Binary

Бинарное представление WKT

Используется в PostGIS, Spatialite

EWKB – WKB, дополненный указанием системы координат

- TWKB – Tiny Well Known Binary

Компактное представление

Поддерживается PostGIS от 2.2.0

Не дружит с GDAL (пока)

Geobuf

- Бинарный формат геоданных
- Основан на Google Protocol Buffers (Pbf), разработан Mapbox
- Конвертируется из/в GeoJSON и TopoJSON
- Обеспечивает частичное чтение
- Очень компактный

MVT – Mapbox Vector Tiles

- Разрабатывается Mapbox с 2013г.
- Поток типа `command,x,y` (идея из SVG)
- Координаты хранятся последовательно в виде дельты от предыдущих
- Координаты в положительных целых числах за счет `zigzag` кодирования
- Хорошо сжимается `gzip` (~1.5 раза)

GeoPackage - универсальный солдат

GeoPackage (gpkg) – шейпфайл модерн
Многолетняя разработка OGC

- Формат базы данных SQLite
- Может хранить еще и растровые данные (тайловые пирамиды аля MBTiles)
- Геометрия в формате WKB
- Несколько слоев в одном файле (и даже несколько наборов тайлов)
- Расширяемый (метаданные!)

Устроим тест!

Возьмем открытые тестовые данные:

- `point` - 315 304 точек

Твиты в Екатеринбурге и окрестностях

- `line` - 28 553 линии (221 285 точек)

Участок дорожной сети Московской области. Данные OSM (стегом highway)

- `poly` – 32 полигона (586 676 точек)

Векторизованное покрытие участка ЦМР ASTER GDEMv2

Покрытие каждого набора - 1x1 градус WGS84

На старт! Внимание! Марш...

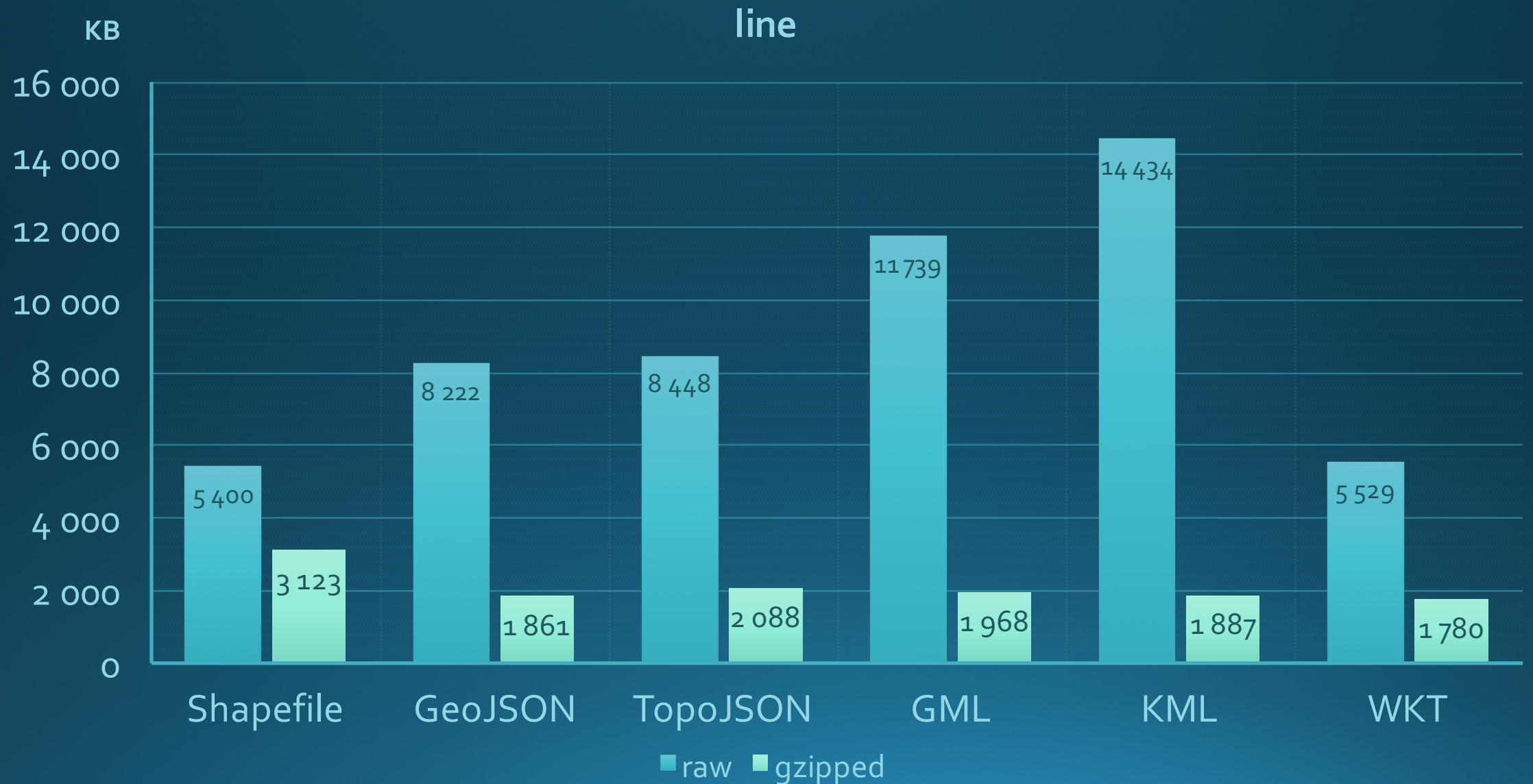
Эталон

	point	line	poly
Shapefile	15 135 kb	5 400 kb	9 431 kb
Shapefile.zip	5 761 kb	3 124 kb	2 720 kb

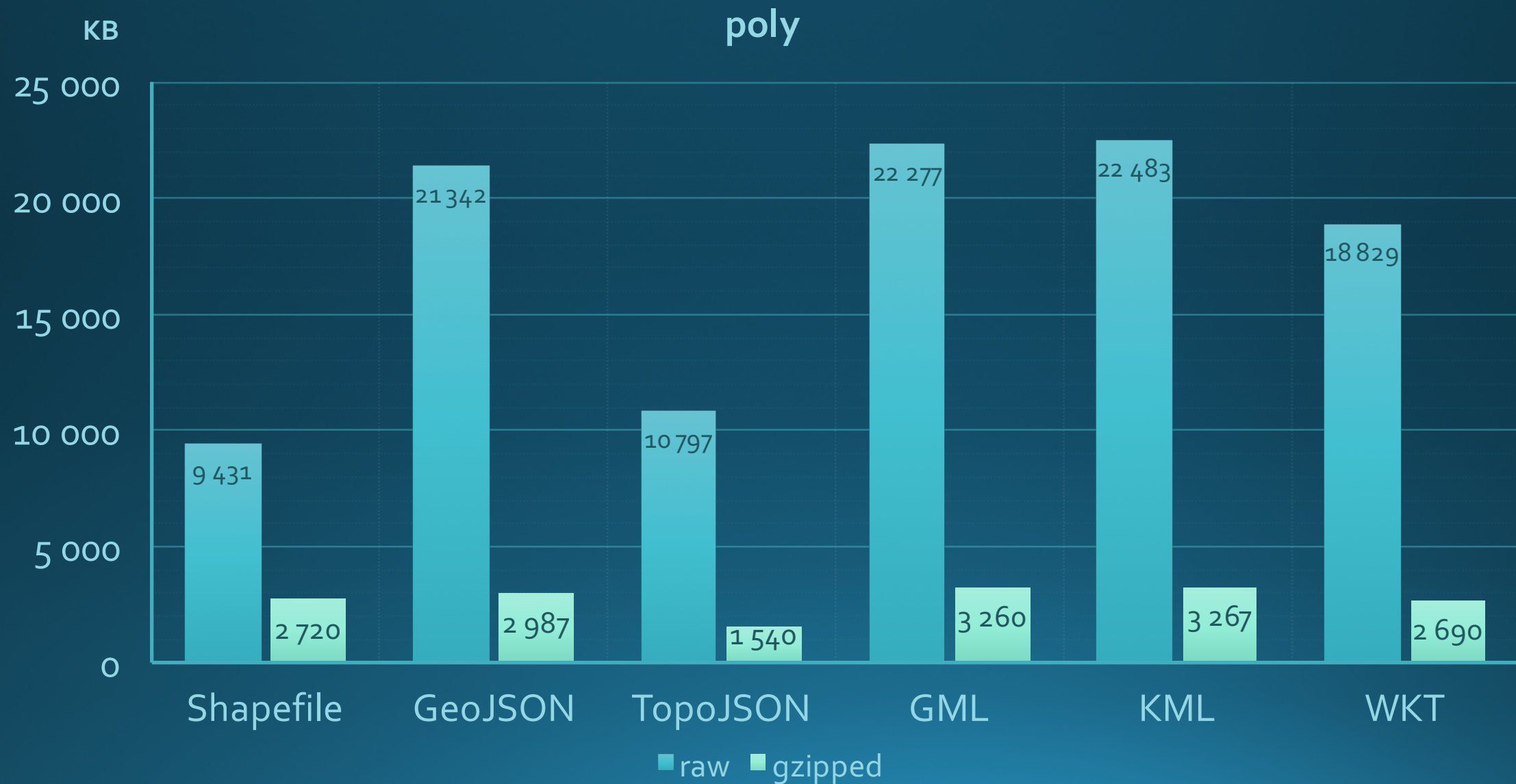
Письменно



Письменно



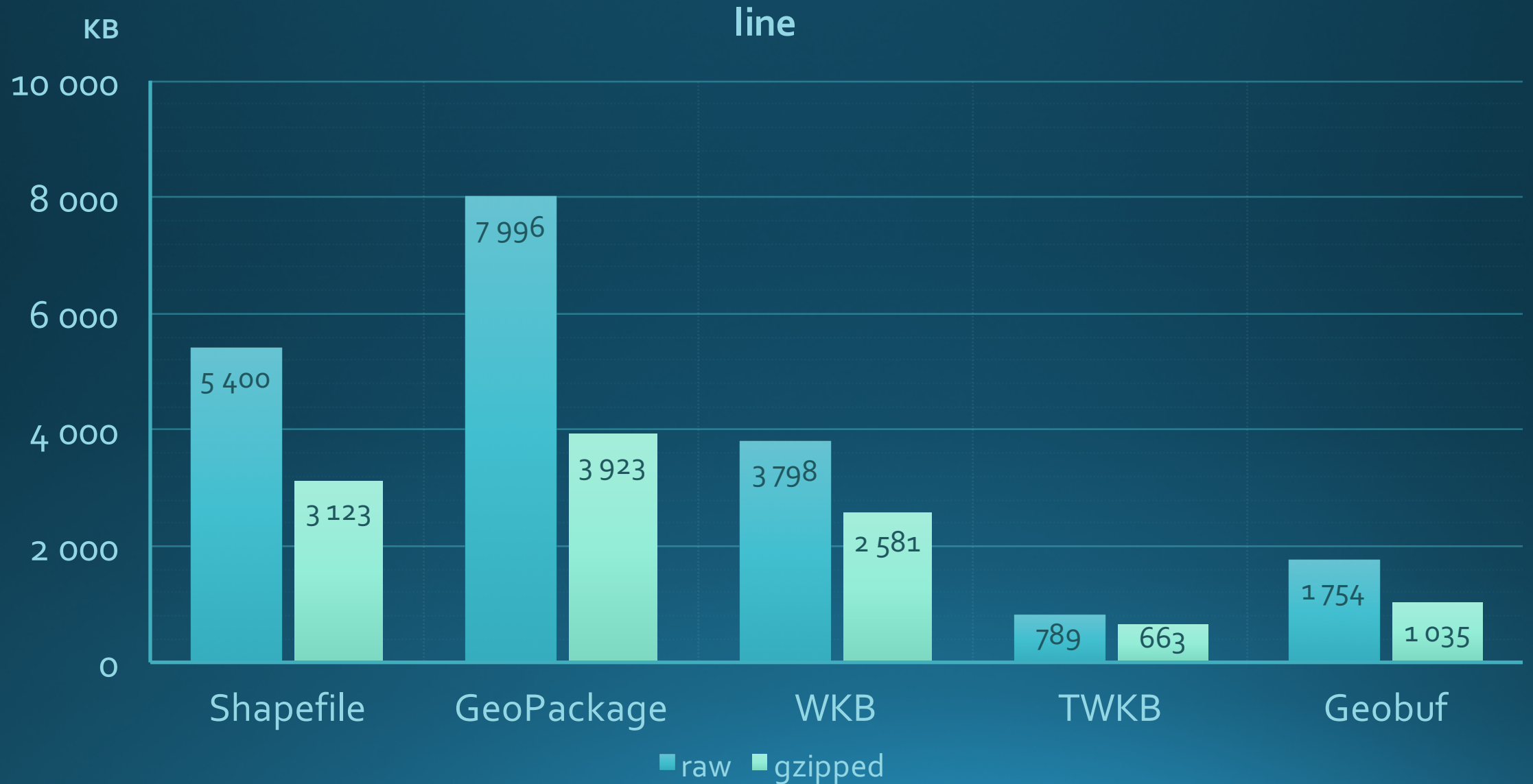
Письменно



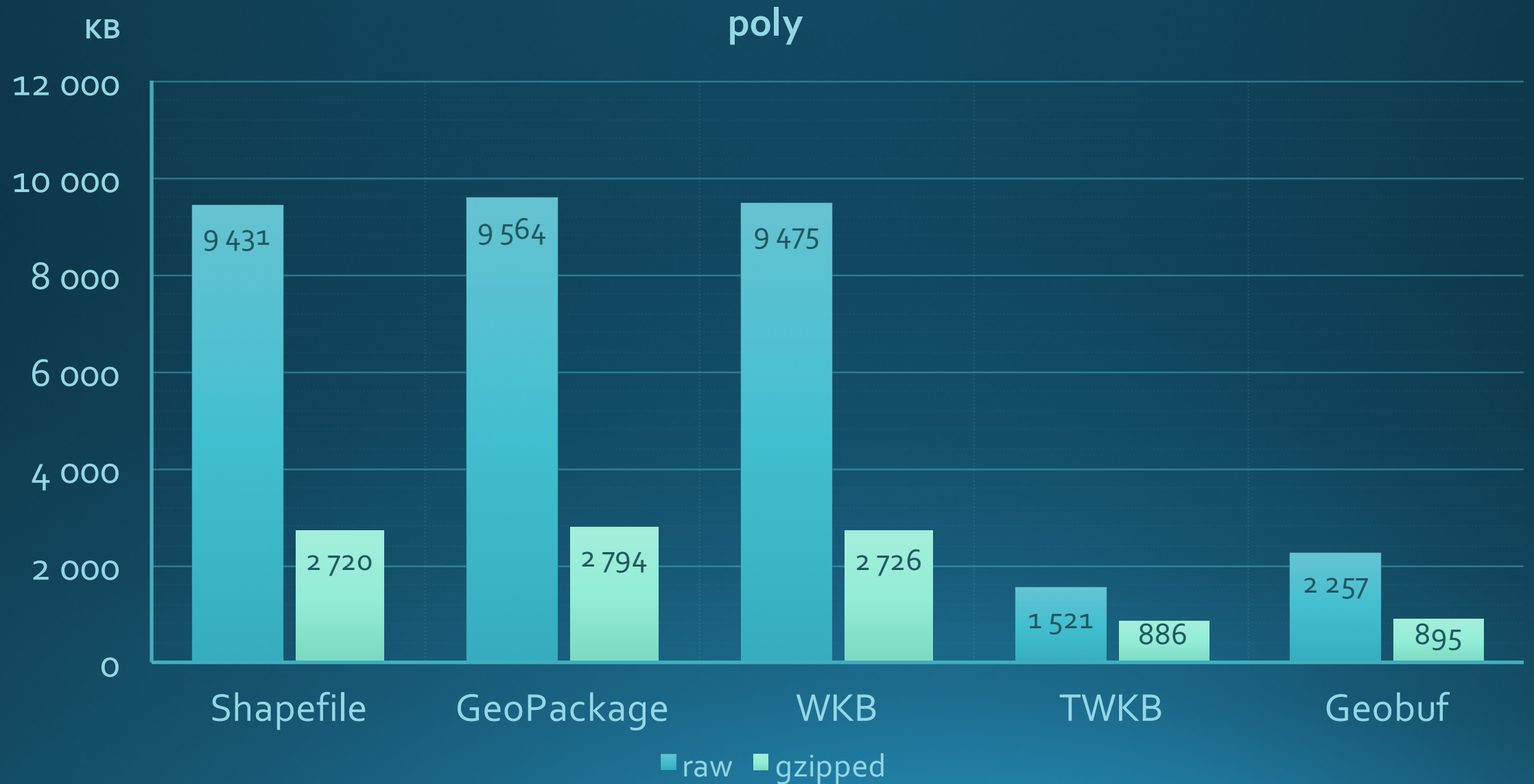
Микромир



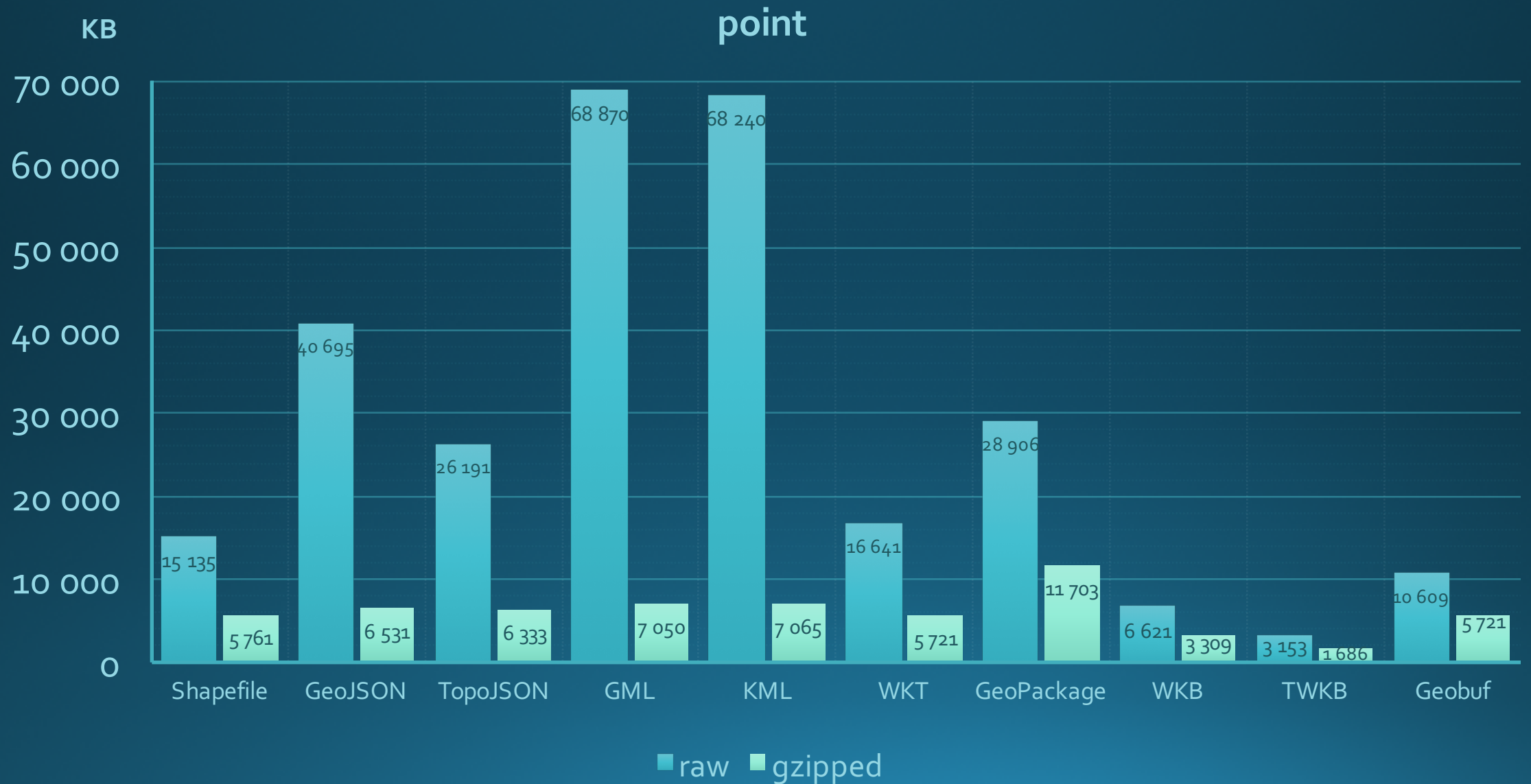
Микромир



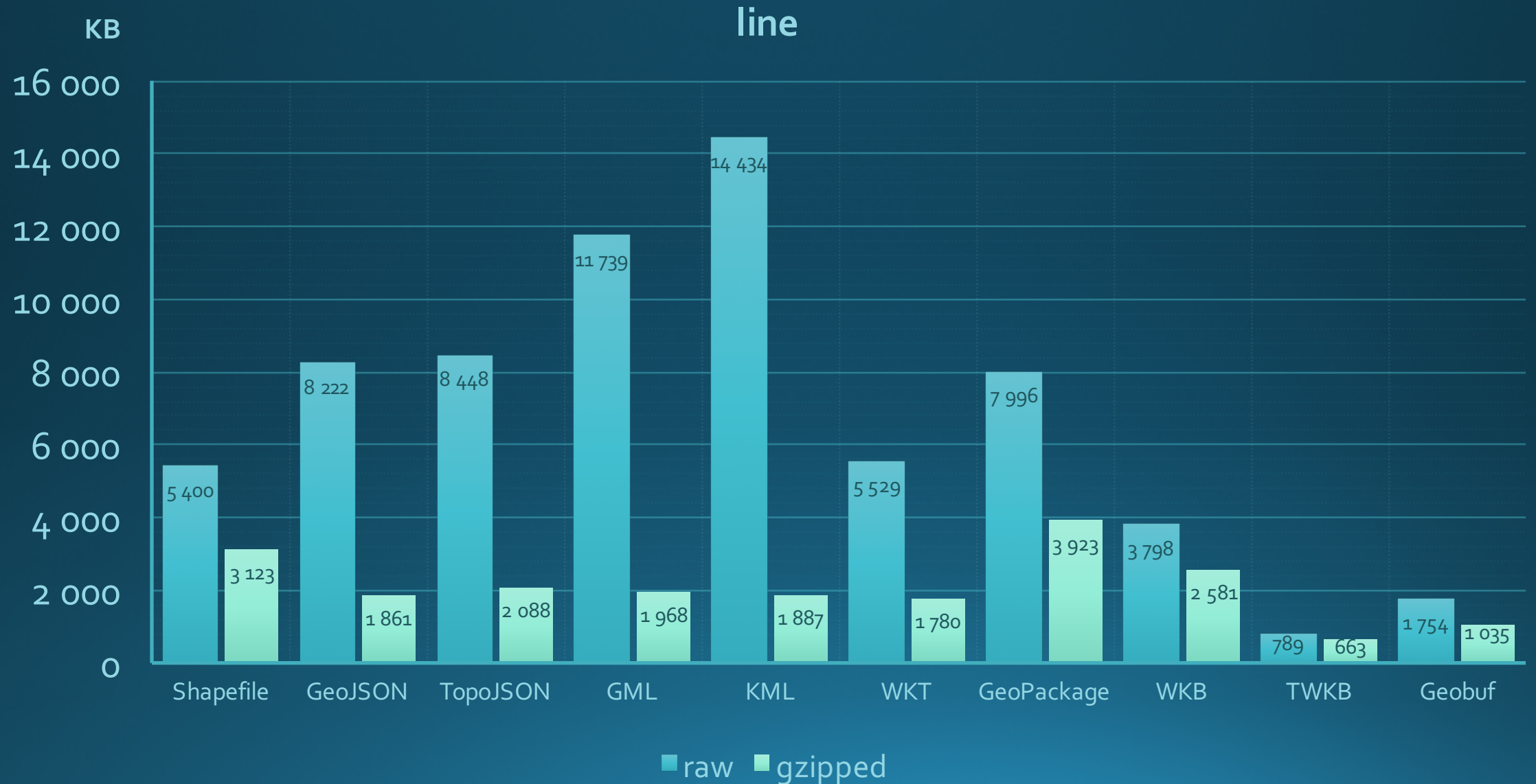
Микромир



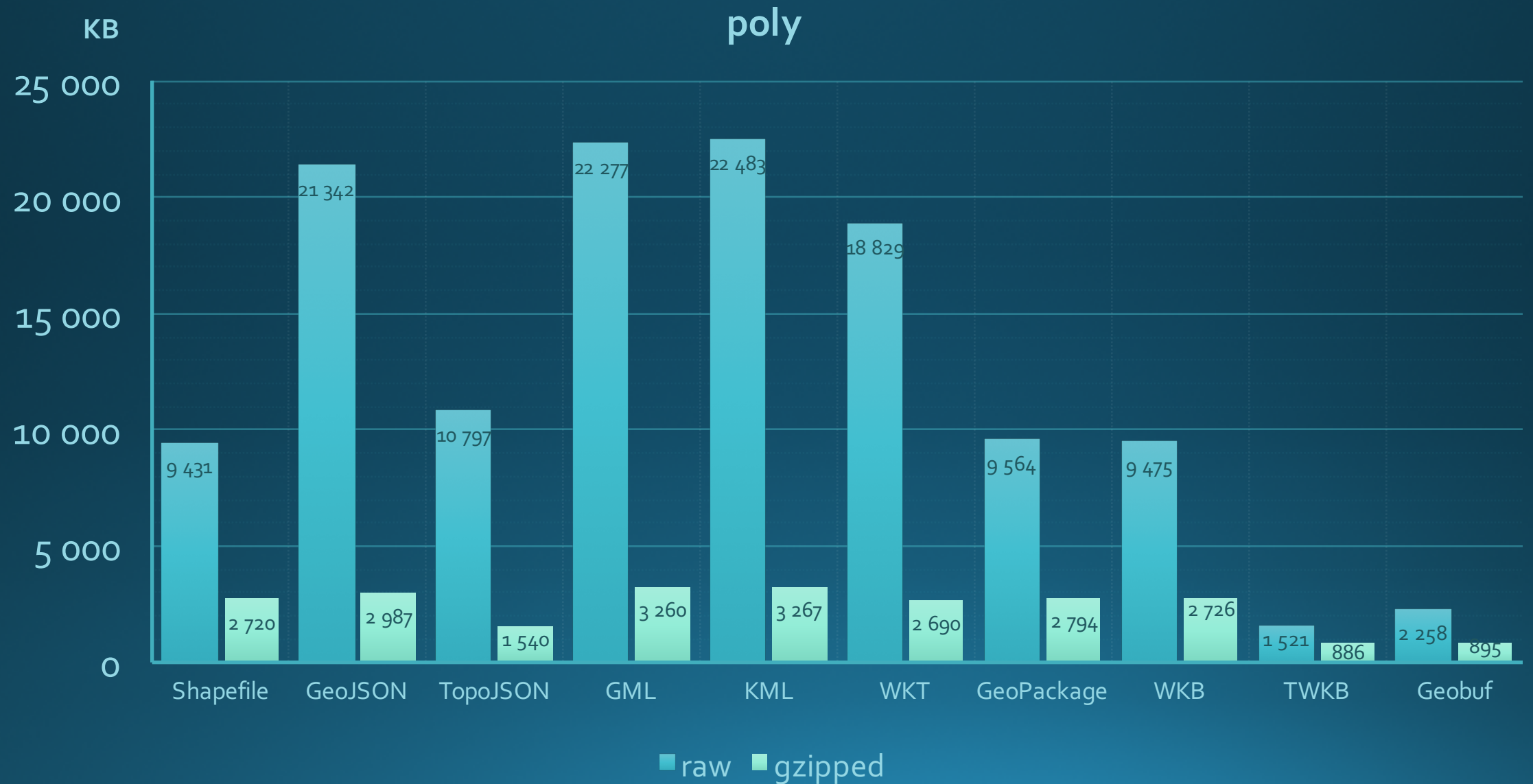
Общий зачет



Общий зачет



Общий зачет



А можно еще меньше?

Go "Lossy"

- Понизить точность координат

Понижение точности до $.00001^\circ$ на тестовых данных:

	point	line	poly
Geobuf / Geobuf.Gz	10609kb / 2876kb	1753kb / 1034kb	2257kb / 895kb
Geobuf / Geobuf.Gz 1e5	6824kb / 2314kb	1230kb / 828 kb	808kb / 477kb

- Упрощение (для линейных и полигональных данных)
Douglas-Peucker, Visvalingam
- Разбить на тайлы (MVT)

Выводы

- Shapefile по-прежнему актуален из-за распространенности, Read/Write поддержки большинством ГИС
- Проблема передачи данных по сети решается компактными форматами – Geobuf, TopoJSON
- Для преимуществ СУБД: WKB (EWKB) / TWKB
- GeoPackage – перспективная эволюция шейпфайла

Светлое будущее

- Генерация Geobuf и MVT в PostGIS
- Драйвер TWKB для GDAL
- Редактирование TopoJSON и Geobuf в QGIS и др.
- Поддержка TWKB в GeoPackage и браузерах (JS)
- Read/Write для GeoPackage в проприетарных ГИС
- Редактирование геоданных в Mapbox GL

ВОПРОСЫ?

Николай Лебедев
nlebedev@gispro.ru
@nextstopsun

