

PGConf.Russia 2023

PosgresPro

**Поместить мир в
обычный ноутбук:
PostgreSQL и
OpenStreetMap**

Анализ геоданных

- Тарификация поездок на такси
- Окружение рядом с коммерческой недвижимостью
- Анализ и планирование инфраструктуры города
- Географические факты и статистика
- Достопримечательности
- Что было бы интересно вам...

OpenStreetMap

Проект, который создаёт и предоставляет свободные географические данные, дает возможность создавать карты всего мира любому человеку, кто этого хочет.

OpenStreetMap и анализ данных

Регулярные выгрузки данных проекта:

<https://planet.openstreetmap.org>

<https://download.geofabrik.de>

Обширная экосистема ПО и способов анализа данных OSM. Популярные программы: osm2pgsql, osmosis, osmium, osm-parquetizer

https://wiki.openstreetmap.org/wiki/Databases_and_data_access_APIs

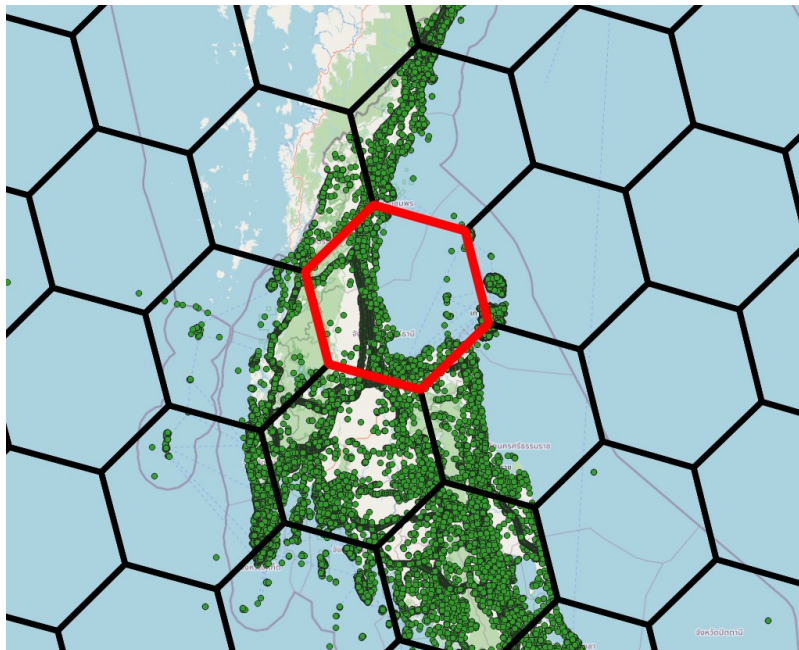
OpenStreetMap H3

Хотелось бы иметь данные в PostgreSQL в формате максимально близком к pgSnapshot, с быстрым механизмом загрузки данных из PBF и возможности параллельной загрузки в БД. Данные секционированы так, чтобы внутри одной секции были объекты расположенные близко друг к другу и сетка разбиения была приблизительно одинаковой по радиусу, как у экватора, так и у полюсов планеты.

https://github.com/igor-suhorukov/openstreetmap_h3



Иерархический гео индекс H3



<https://h3geo.org>

| Res | Число ячеек |
|----------|------------------|
| 0 | 122 |
| 1 | 842 |
| 2 | 5882 |
| 3 | 41162 |
| 4 | 288122 |
| 5 | 2016842 |
| 6 | 14117882 |
| 7 | 98825162 |
| 8 | 691776122 |
| 9 | 4842432842 |
| 10 | 33897029882 |
| 15 | 569707381193162 |

PostGIS+PG15+Citus+H3=Docker

```
cd openstreetmap_h3/postgis_docker-master/Dockerfile
docker build -t postgres15_postgis .
```

```
FROM postgis/postgis:15-3.3
RUN apt-get update
RUN apt-get install -y pgxnclient make pip git postgresql-server-dev-all clang-format clang-tidy g++ zstd
RUN apt-get update && apt-get install -y python3-requests postgresql-plpython3-${PG_MAJOR} apt-utils
RUN pip install cmake --upgrade
RUN pgxn install h3

RUN apt-get update && apt-get install -y parallel

ADD docker-entrypoint-initdb.d /docker-entrypoint-initdb.d

RUN apt-get install -y curl
RUN curl https://install.citusdata.com/community/deb.sh | bash
RUN apt-get purge -y --auto-remove curl
RUN apt-get -y install apt-utils postgresql-15-citus-11.2
RUN echo "shared_preload_libraries='citus'" >> /usr/share/postgresql/postgresql.conf.sample
RUN echo "shared_preload_libraries='citus'" >> /var/lib/postgresql/data/postgresql.conf
```

Весь мир из planet.pbf

```
docker start postgis-planet-220704
```

```
psql -h 127.0.0.1 -p 5432 -U postgres -d osmworld
```

| Table | Rows | Total Size | Table Size | Index(es) Size | TOAST Size |
|---------------|------|------------------|------------------|-----------------|-----------------|
| *** TOTAL *** | ~1B | 587 GB (100.00%) | 501 GB (100.00%) | 54 GB (100.00%) | 32 GB (100.00%) |

Широка страна моя родная

```
~/dev/map/russia$ wget https://download.geofabrik.de/russia-latest.osm.pbf
```

```
git clone https://github.com/igor-suhorukov/openstreetmap_h3.git
```

```
cd openstreetmap_h3 && mvn package
```

```
java -jar target/osm-to-pgsnapshot-schema-ng-1.0-SNAPSHOT.jar -source_pbf ~/dev/map/russia/russia-latest.osm.pbf
```

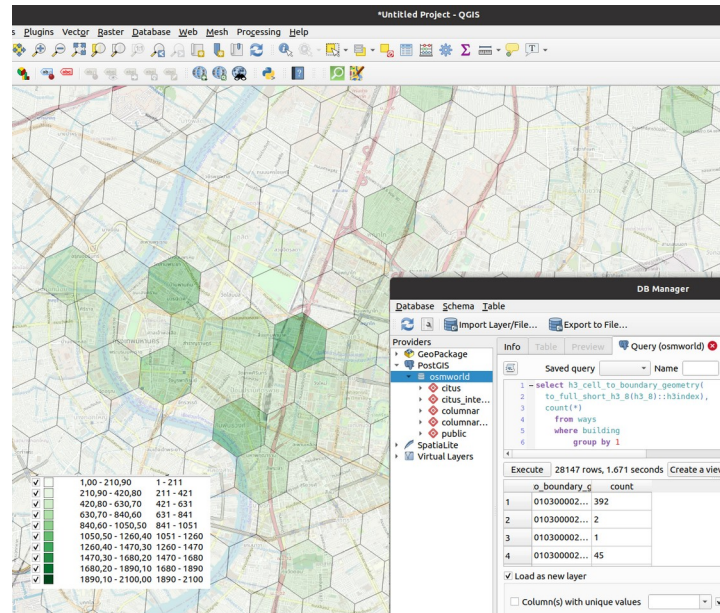
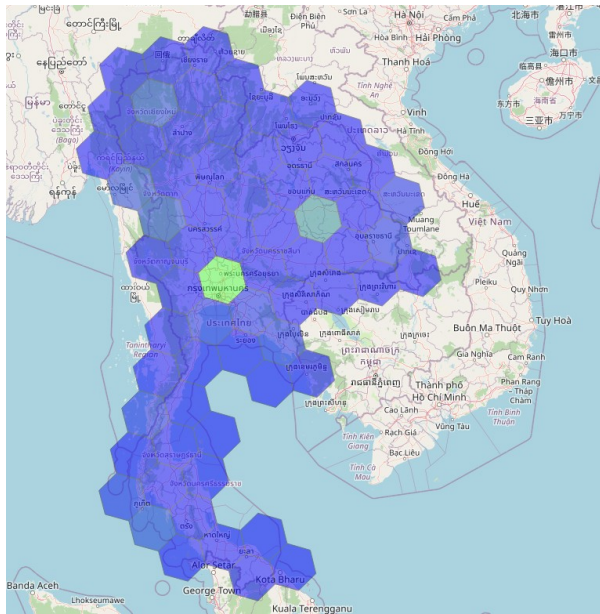
Протяженность дорог

```
osmworld=# select sum(ST_Length(linestring::geography))
           from ways
           where tags->'highway' is not null
           and tags->'highway' not in ('platform','bus_stop','corridor','rest_area');
           sum
-----
4265048225.969036
(1 row)

Time: 40327,153 ms (00:40,327)
```

Примеры на ноутбуке

Вы можете повторить все рассказанное мной на данных OSM в PostGIS



The screenshot shows a QGIS interface. On the left, a map displays a network of roads. A specific road segment is highlighted in red. On the right, the 'Query Results' panel is open, displaying a table of feature attributes for the selected road segment.

| Feature | Value |
|--------------------|--|
| QueryLayer_3 | |
| uid | 17154 |
| (Derived) | |
| (Actions) | |
| uid | 17154 |
| closed | false |
| building | false |
| highway | true |
| h3_3 | 25648 |
| h3_8 | 1681213729 |
| scale | 15,738768 |
| tags | highway: service |
| bbox | SRID=4326;POLYGON((99.3345319 9.117873900... |
| centre | SRID=4326;POINT(99.33458725 9.11800595) |
| id | 935774700 |
| points | 8671540288, 8671540287 |
| h3_8_regions | NULL |
| h3_3_multi_regions | NULL |

Citus: Сравним Columnar с Heap

```
java -jar target/osm-to-pgsnapshot-schema-ng-1.0-SNAPSHOT.jar -columnar_storage -  
source_pbf ~/dev/map/russia/russia-latest.osm.pbf
```

```
osmworld(heap)=# select h3_3, count(*) from ways where building group by 1 order by  
2 desc limit 15;
```

Time: **12787,813** ms (00:12,788)

```
osmworld(columnar)=# select h3_3, count(*) from ways where building group by 1  
order by 2 desc limit 15;
```

Time: **1212,481** ms (00:01,212) **10X быстрее**

OpenStreetMap H3 → Apache Parquet

Если нужно обрабатывать еще большие объемы геоданных на основе OpenStreetMap, обучать ML модели, выполнять огромные объемы вычислений над данными в Apache Spark и Apache Sedona, загружать данные в кластер ClickHouse, предобрабатывать в DuckDB Geo, то это все возможно с помощью опций `-arrow_format PARQUET` (или `ARROW_IPC`)

https://github.com/igor-suhorukov/openstreetmap_h3



PGConf.Russia 2023

https://github.com/igor-suhorukov/openstreetmap_h3

<https://postgis.net>

<https://h3geo.org>

PostgresPro



PGConf.Russia 2023

PosgresPro

Q & A

PGConf.Russia 2023

PosgresPro

Спасибо!