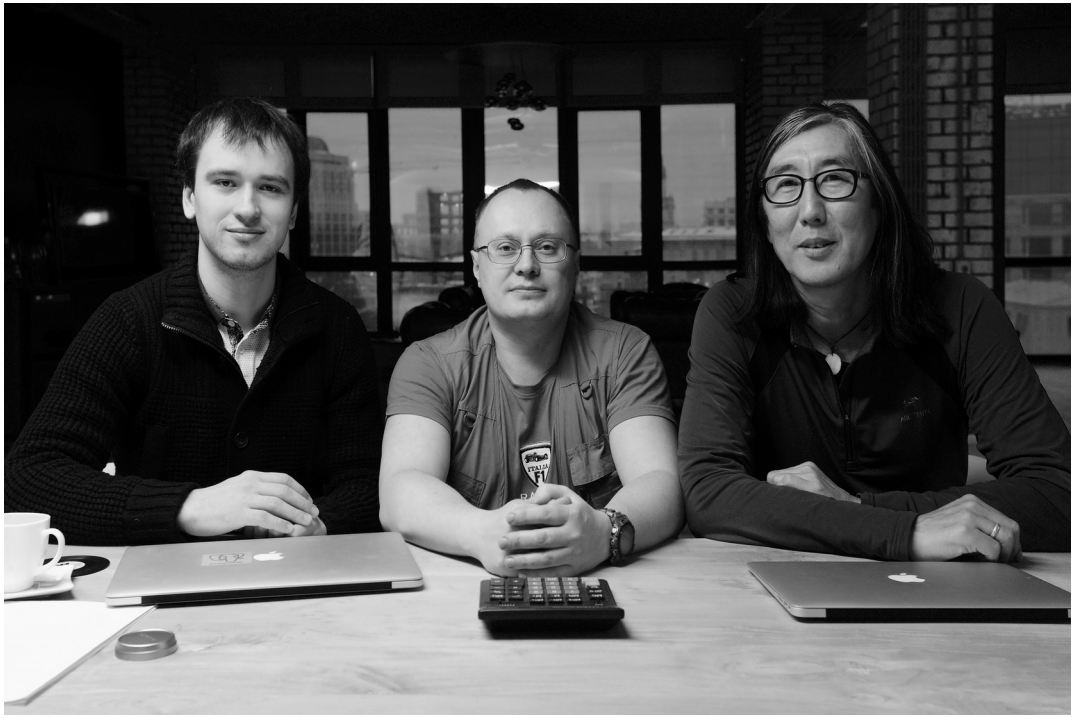
Spatial indexes in PostgreSQL for astronomy

Alexander Korotkov[1], Oleg Bartunov[12]

[1]Postgres Professional

[2]SAI MSU

Alexander Korotkov, Teodor Sigaev, Oleg Bartunov



- Major contributors to PostgreSQL
- Co-founders of Postgres Professional

PostgreSQL CORE
- Locale support
- PostgreSQL extendability:
- GiST(KNN), GIN, SP-GiST
- Full Text Search (FTS)
- NoSQL (hstore, jsonb)
- Indexed regexp search
- Custom AM & Generic WAL
- Pluggable table engines (WIP)

Extensions:
- Intarray
- Pg_trgm
- Ltree
- Hstore
- plantuner

Why PostgreSQL?

- Mature DBMS with all necessaries
- Open Source product distibuted under BSD-like license
- Big and responsible community
- Outstanding extendability
- Rich features to work with

WSDB Whole Sky DataBase

- Database of astronomic catalogues in Cambridge University
- ~5 dbs,~ 40 users, up to ~ 10^7 queries per day, size 40Tb
- pg 9.4 + q3c + hstore
- Example of research: Koposov, S. E., Belokurov, V., Torrealba, G., & Evans, N. W. (2015). Beasts of the Southern Wild: Discovery of nine Ultra Faint satellites in the vicinity of the Magellanic Clouds. The Astrophysical Journal, 805(2), 130.

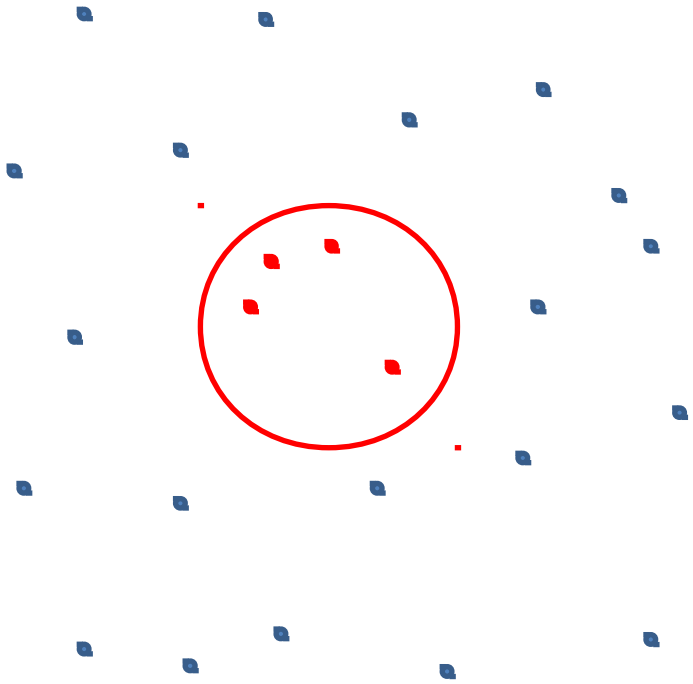
Gaia Alerts Database

- Real time Detection of alerts in the Gaia
- ~10 dbs, 10 users, up to ~ 10^6 queries per day, size 30Tb
- pg 9.3 + synchronous replication + q3c
- Example of research: Campbell, H. C., Marsh, T. R., Fraser, M., Hodgkin, S. T., de Miguel, E., Gänsicke, B. T., ... & Koposov, S. E. (2015). Total eclipse of the heart: the AM CVn Gaia14aae/ASSASN-14cn. Monthly Notices of the Royal Astronomical Society, 452(1), 1060-1067.

MASTER database

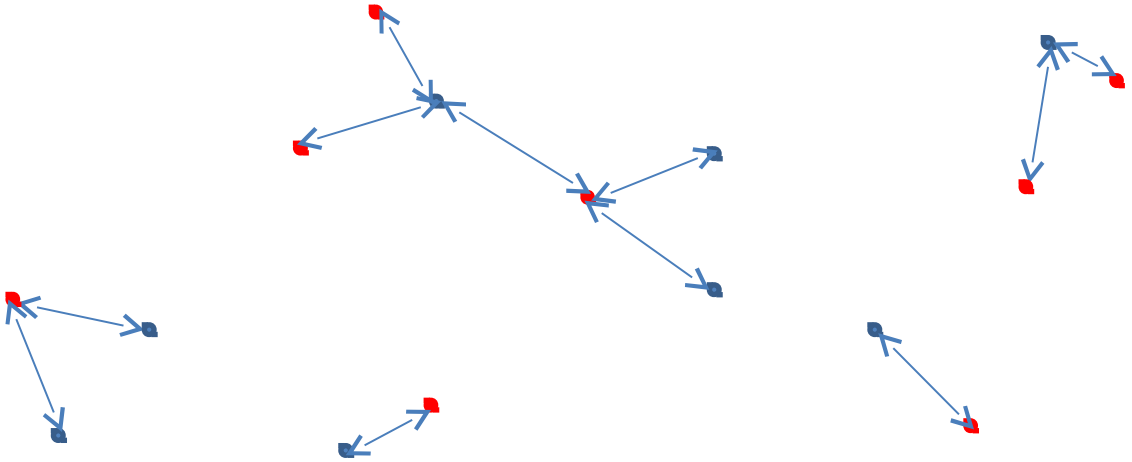
- Robotic net of telescopes by SAI MSU
- 8 observatories (5 in Russia, 3 outside)
- total size ~100TB
- pg 9.0-9.4 + pgsphere + replication
- See:
  - Lipunov, Vladimir, et al. "Master robotic net." Advances in Astronomy 2010 (2010).
  - Kornilov, Victor G., et al. "Robotic optical telescopes global network MASTER II. Equipment, structure, algorithms." Experimental Astronomy 33.1 (2012): 173-196.

Radial query



PROFESSIONAL Postgres

Crossmatch

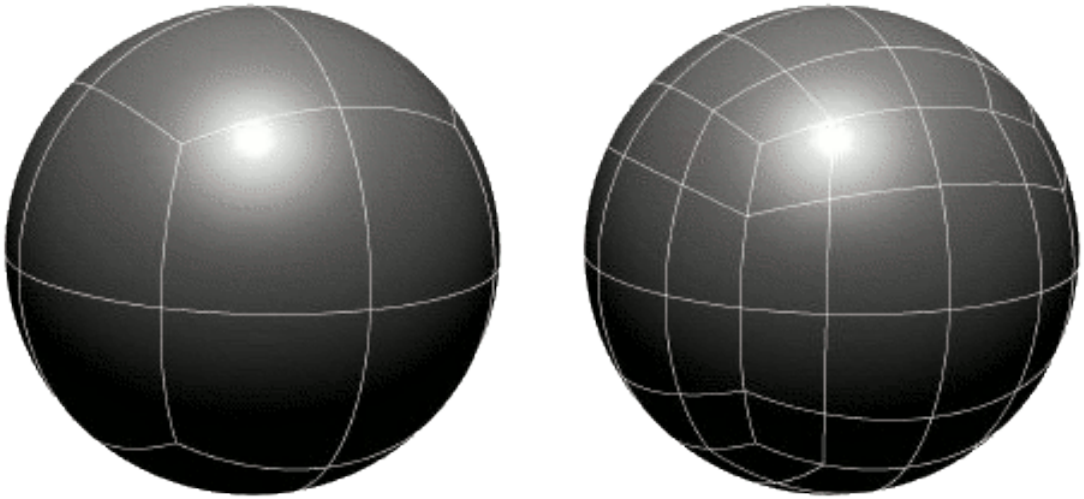


Indexing methods

- haversine (haive)
- Q3C
- pgSphere
- PostGIS

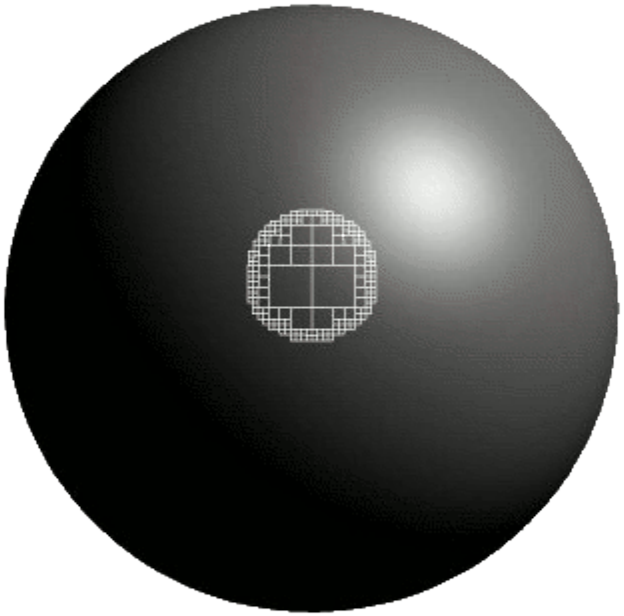
q3c structure

- Cube is inscribed into sphere
- Central projection of cube into sphere
- Quad-tree on cube faces transforms spherical coordinates into integer (IPIX)



PROFESSIONAL Postgres

Querying using q3c



PROFESSIONAL Postgres

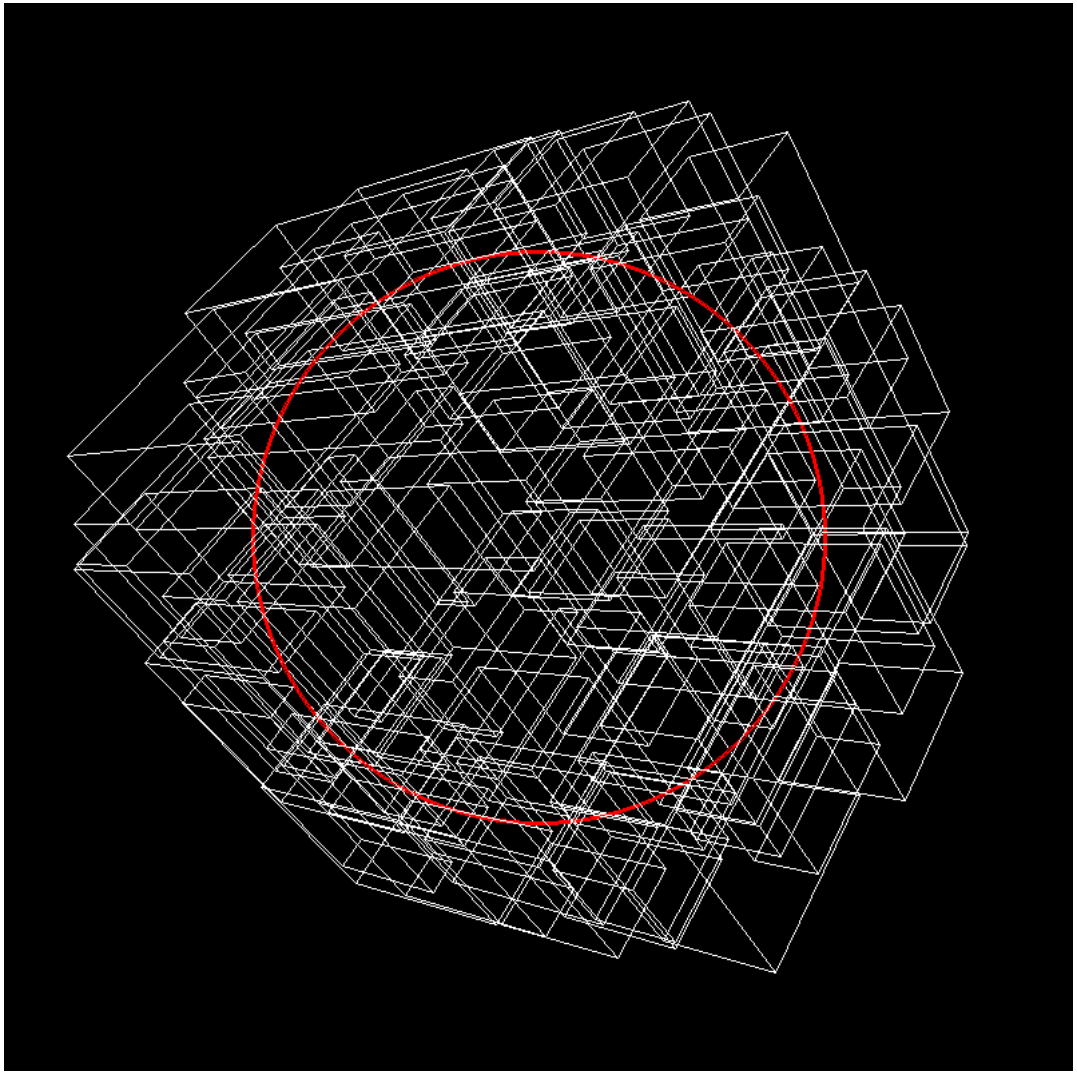
R-tree

Generalized search tree (GiST)

- Generalization over R-tree and its variances
- PostgreSQL GiST is the only full featured implementation of GiST
- A lot of applications (operator classes) to various tasks

PROFESSIONAL Postgres

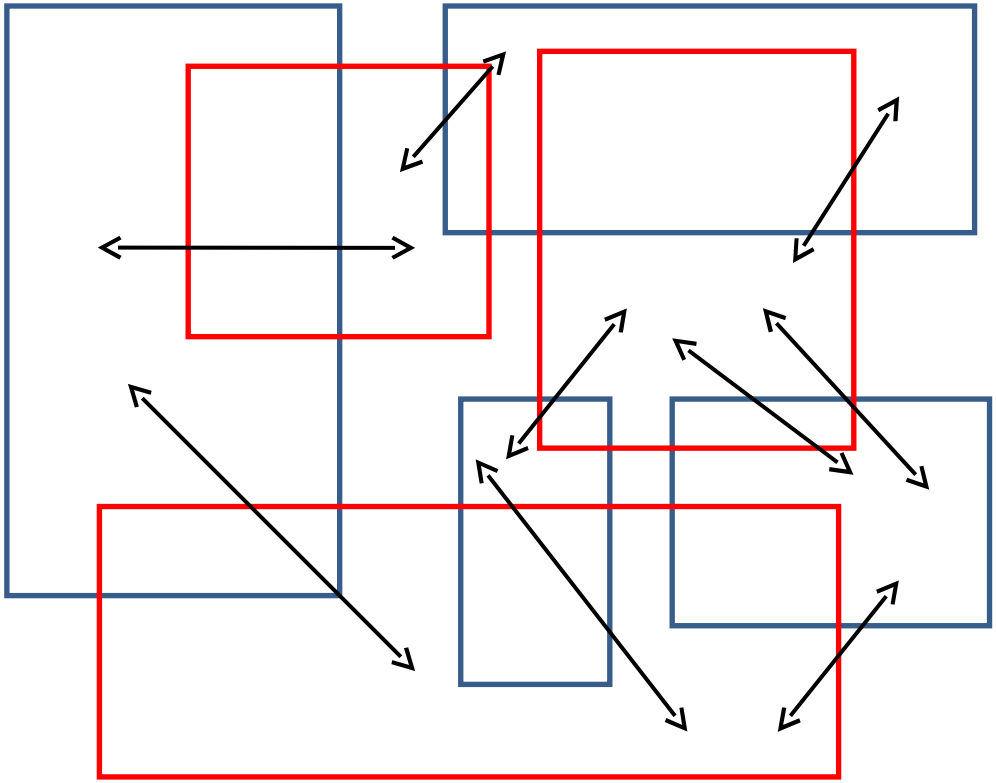
pgsphere



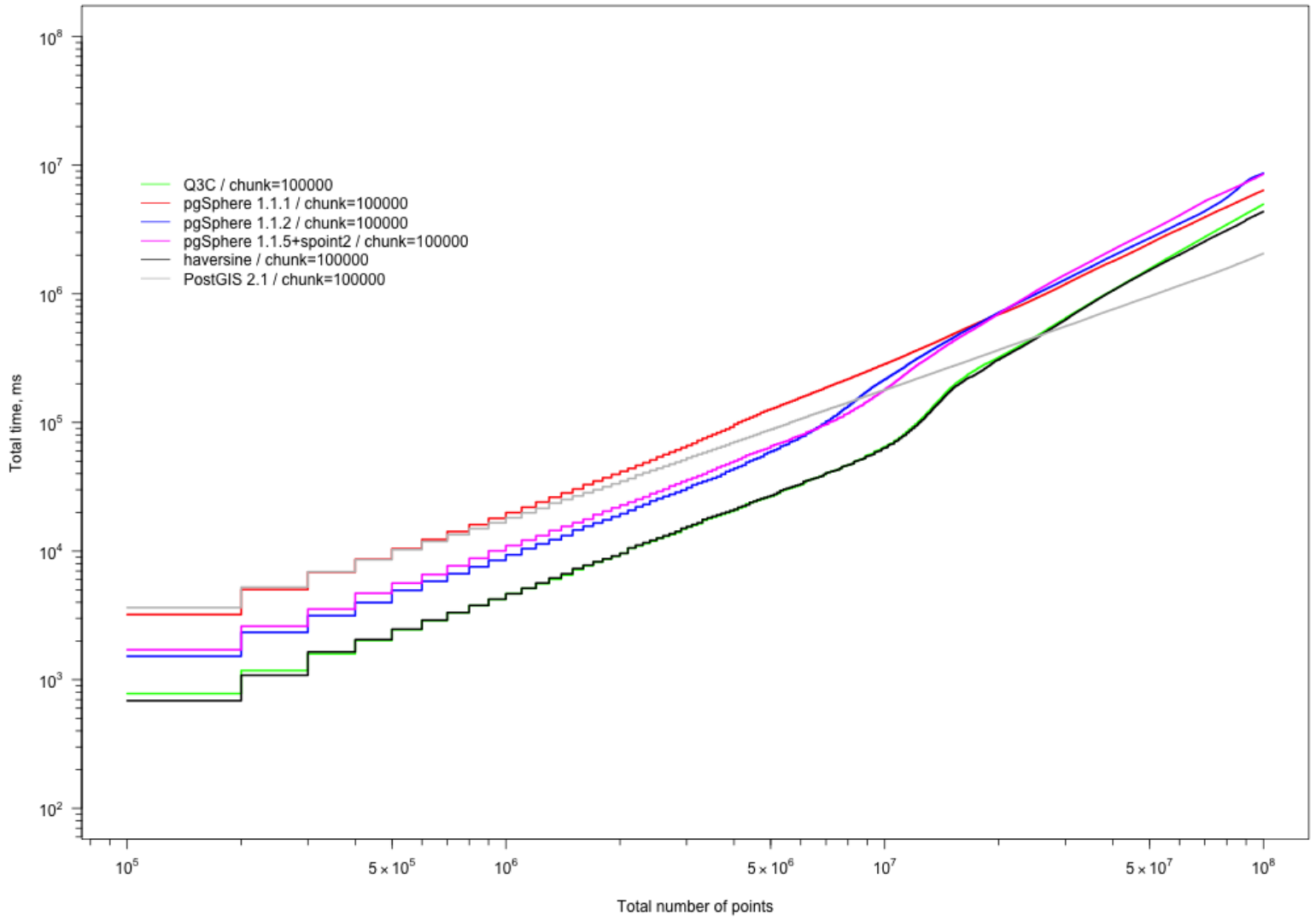
pgsphere versions

- pgSphere 1.1.1 – last version before this research
- pgSphere 1.1.2 – bug fixes, new node splitting algorithm
- pgSphere 1.1.5 – new operator class spoint2, which able to store source points in index leafs

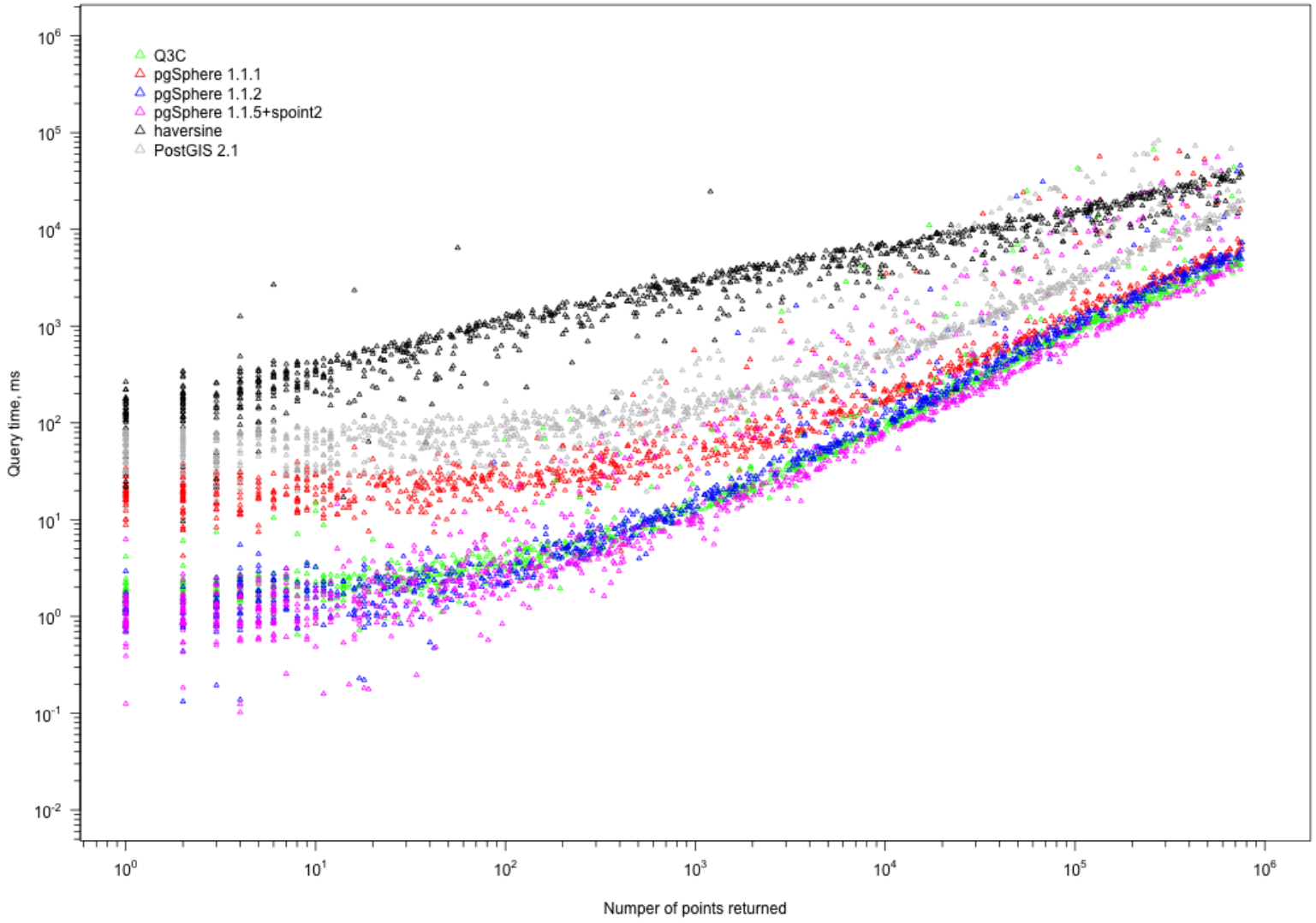
Crossmatch by two indexes



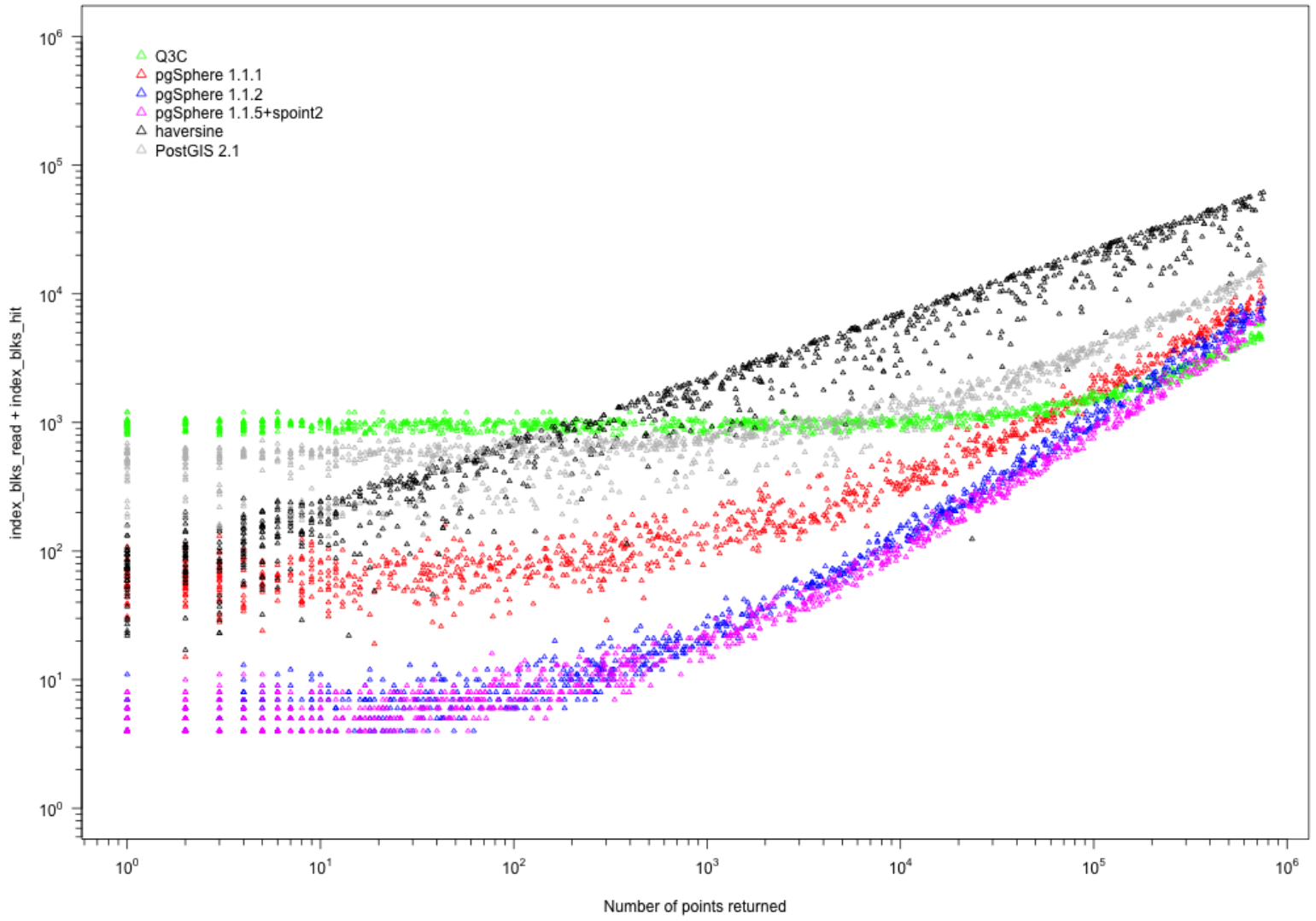
Time required for insertion



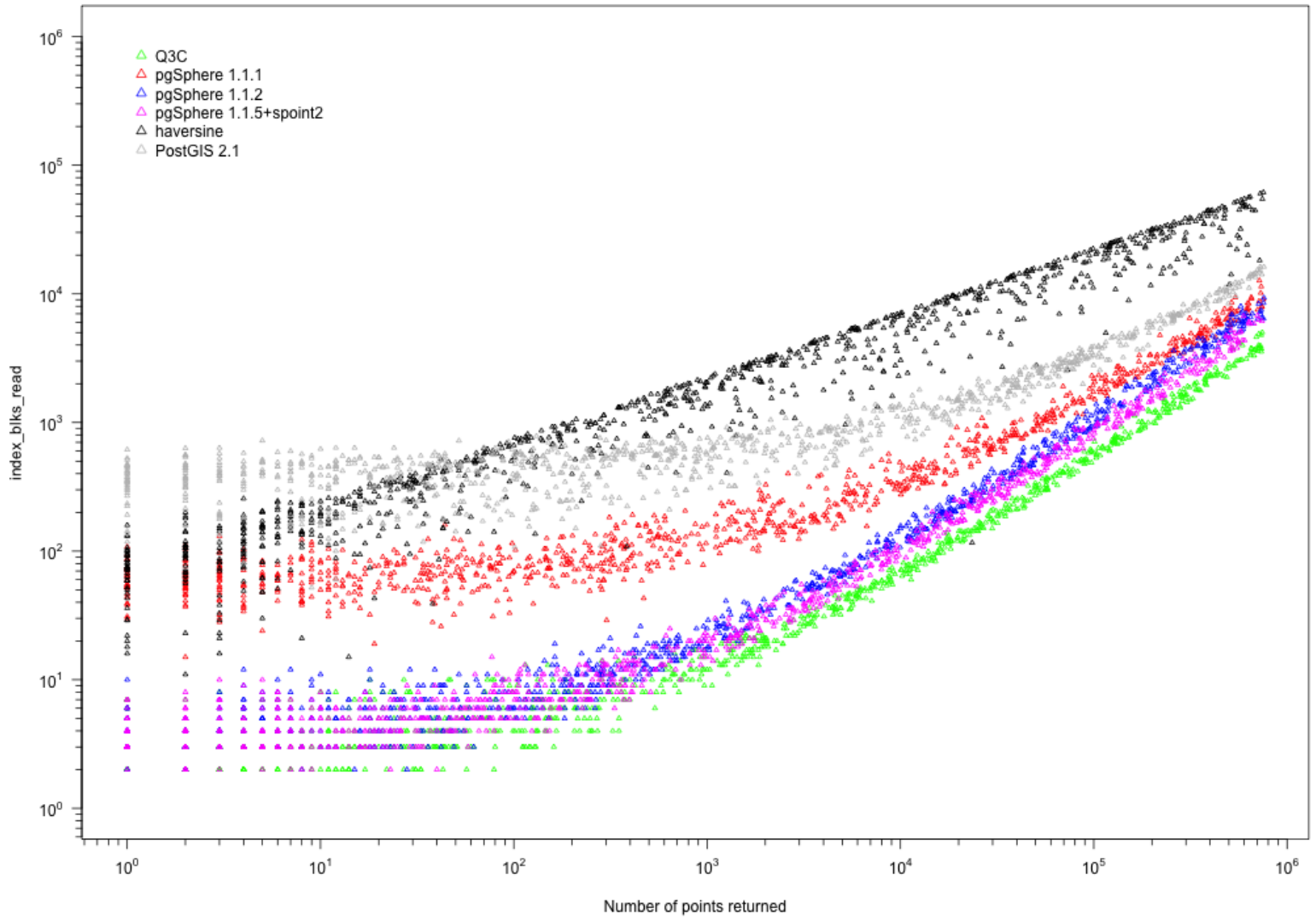
Time required for radial queries



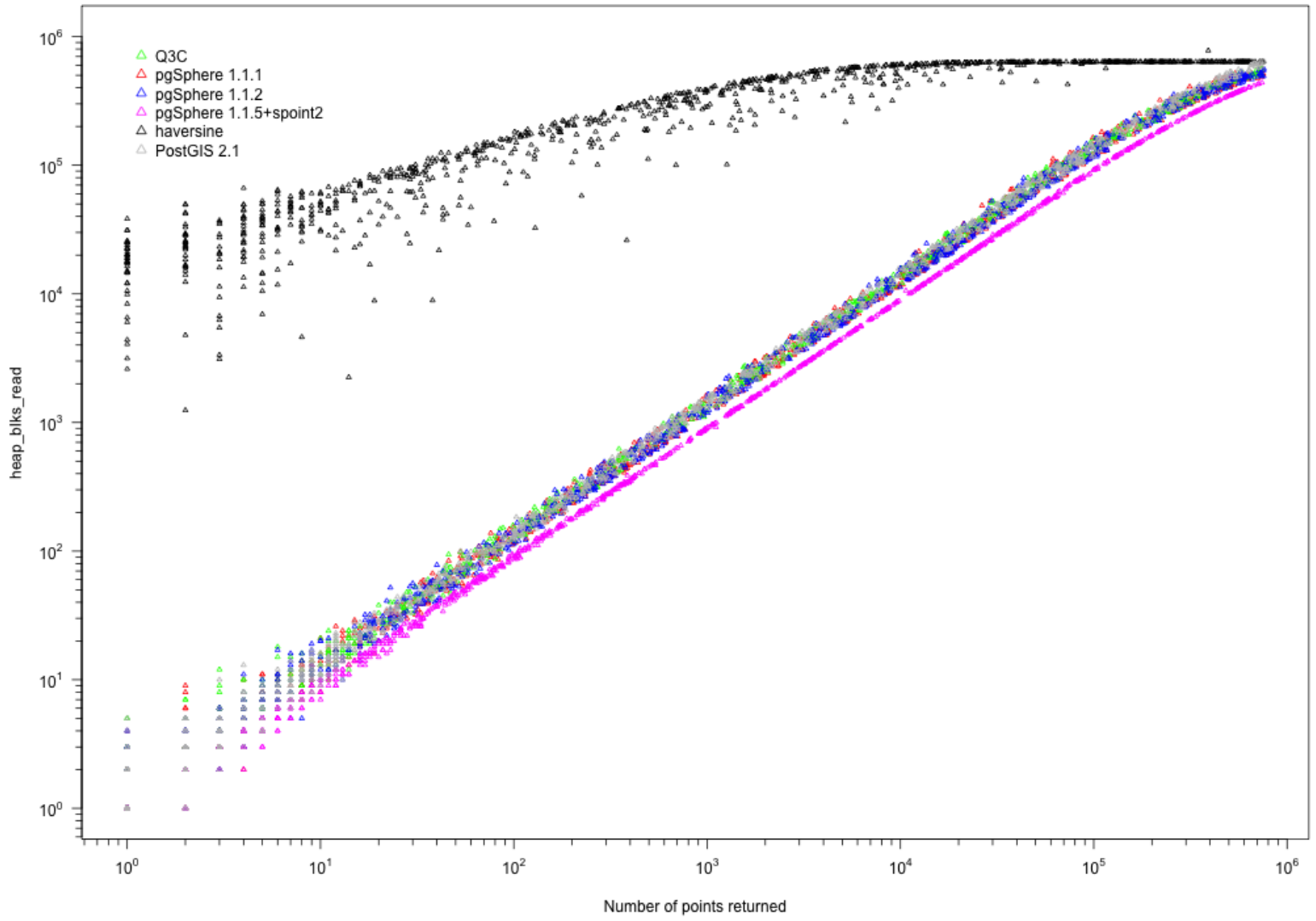
Number of used index blocks



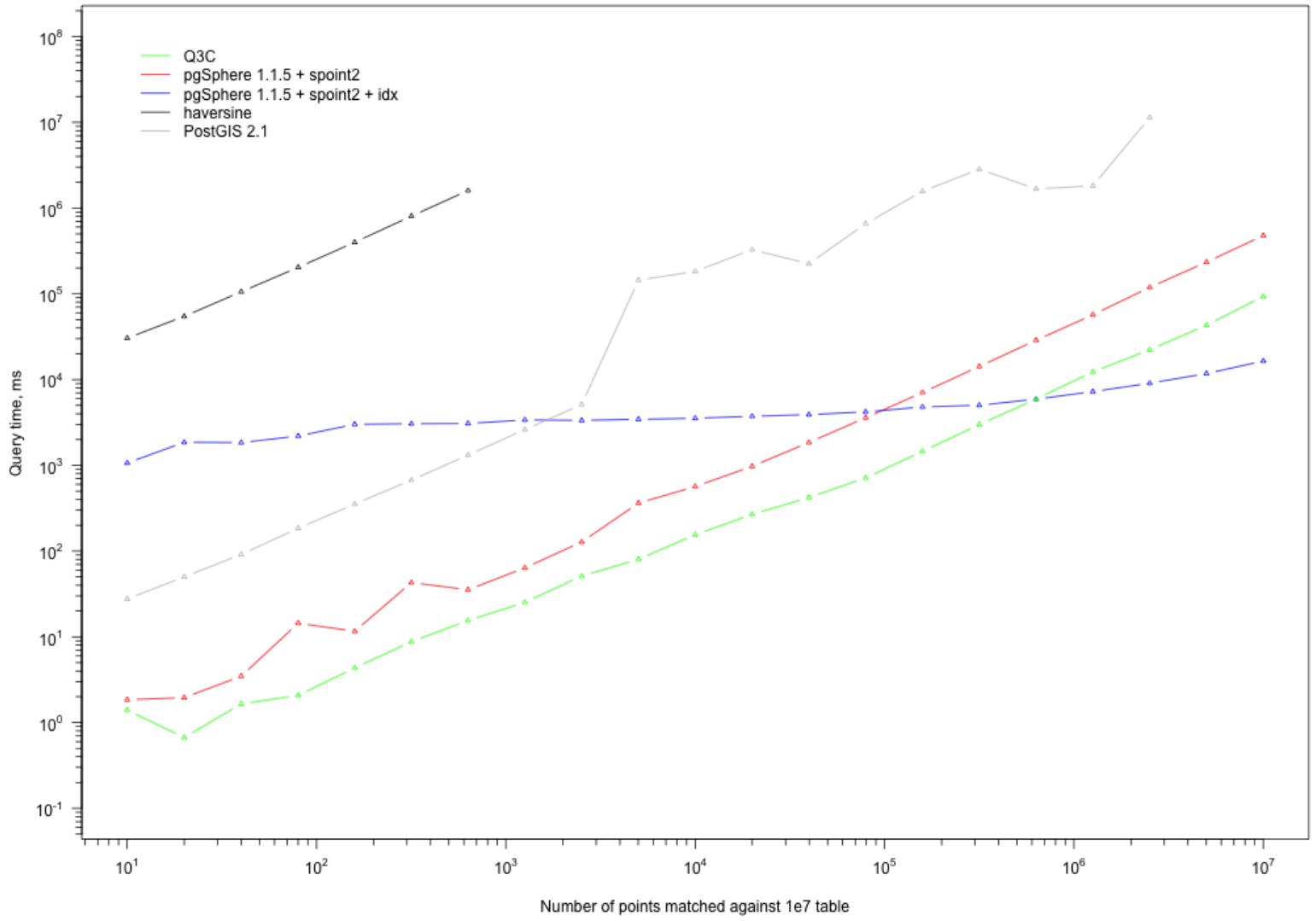
Number of read index blocks with shared_buffers = 512 МБ



Number of read heap blocks

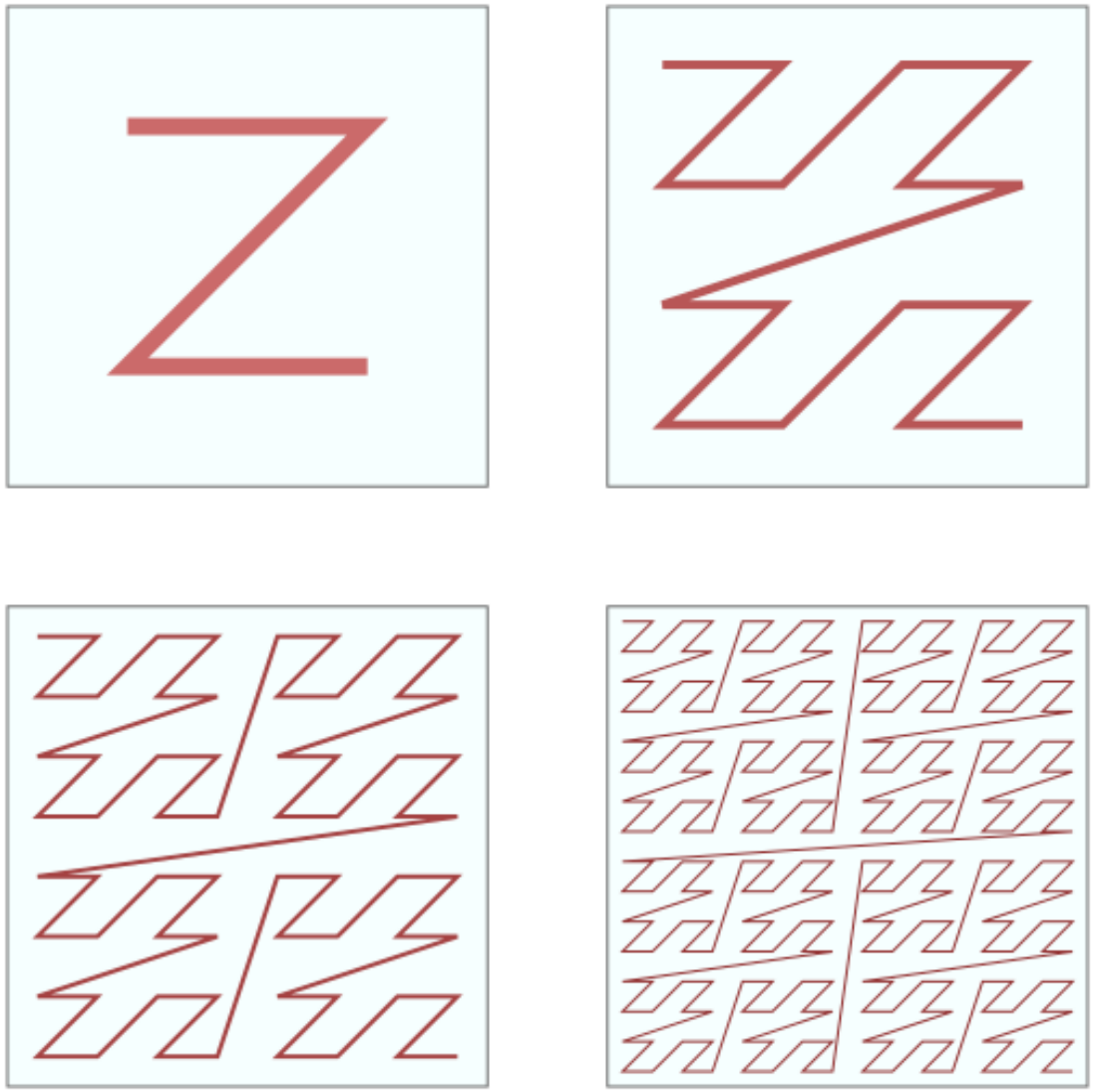


Time of crossmatch execution



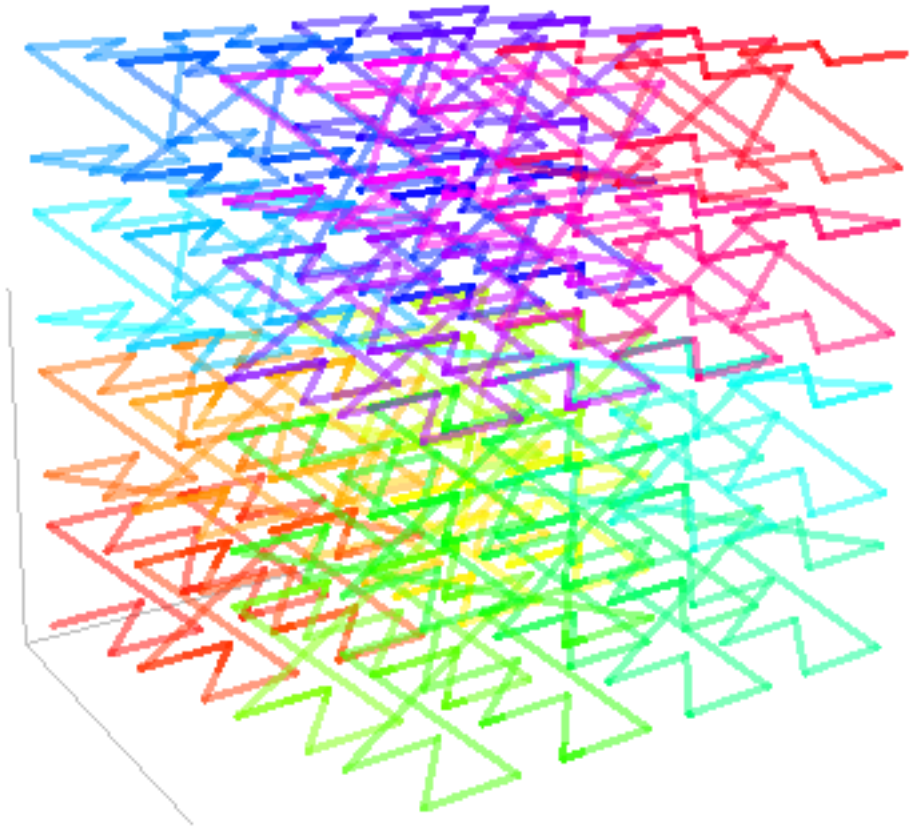
Postgres PROFESSIONAL

Z-curve



Postgres PROFESSIONAL

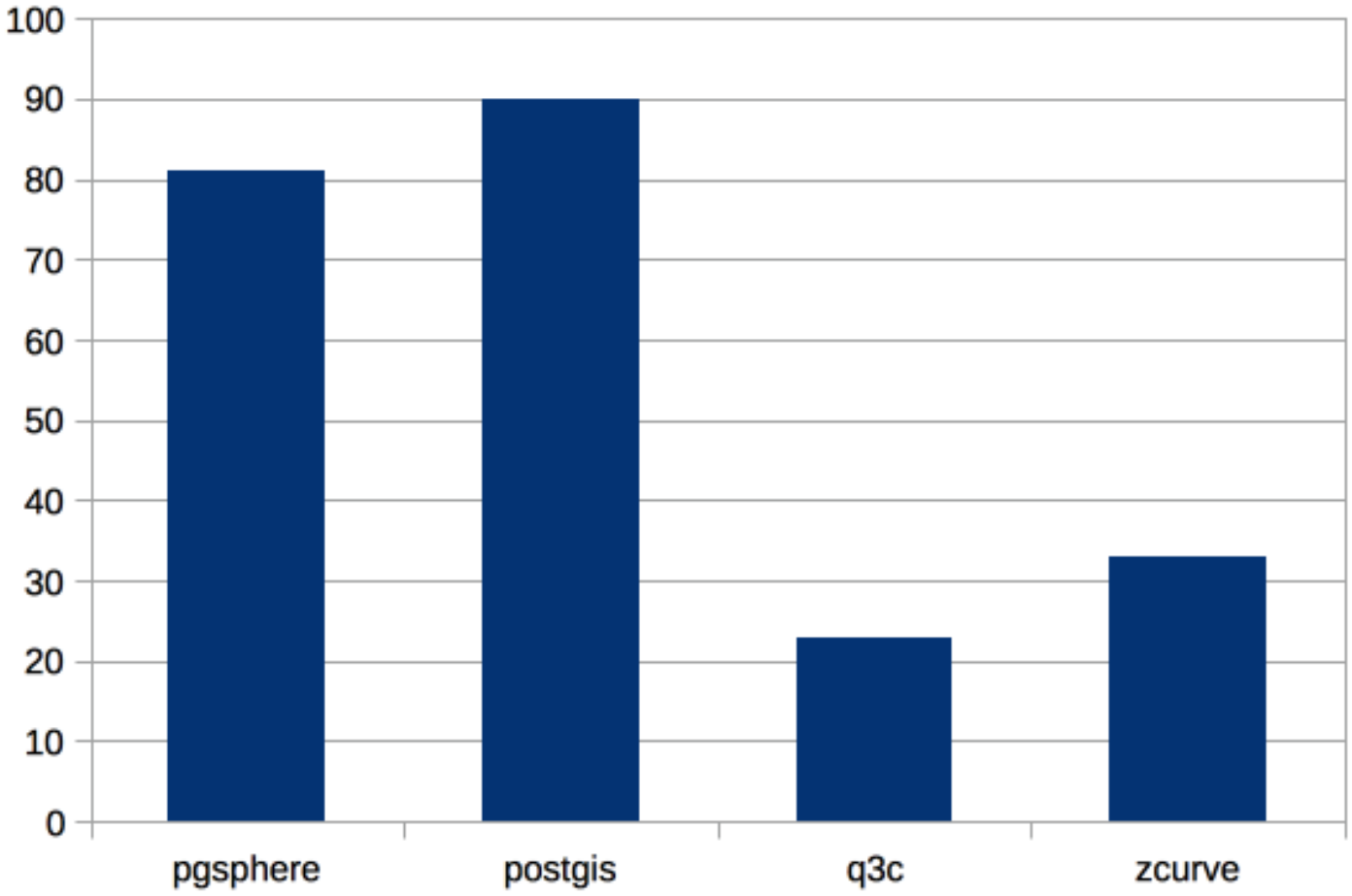
3d-zcurve



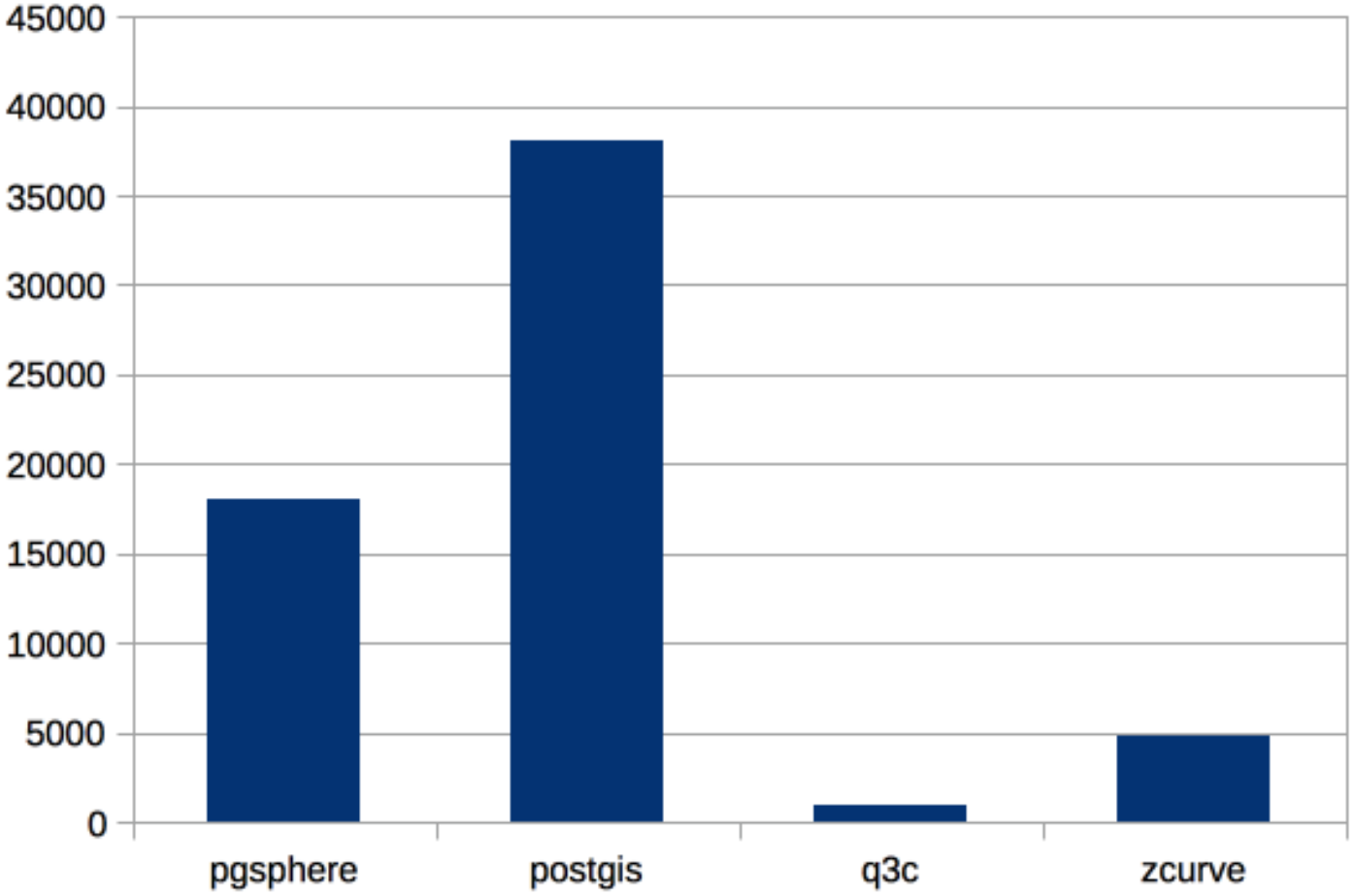
Experiments on real-life dataset

- Nomad catalogue
- 10^9 points
- Size of table without indexes = 182 GB

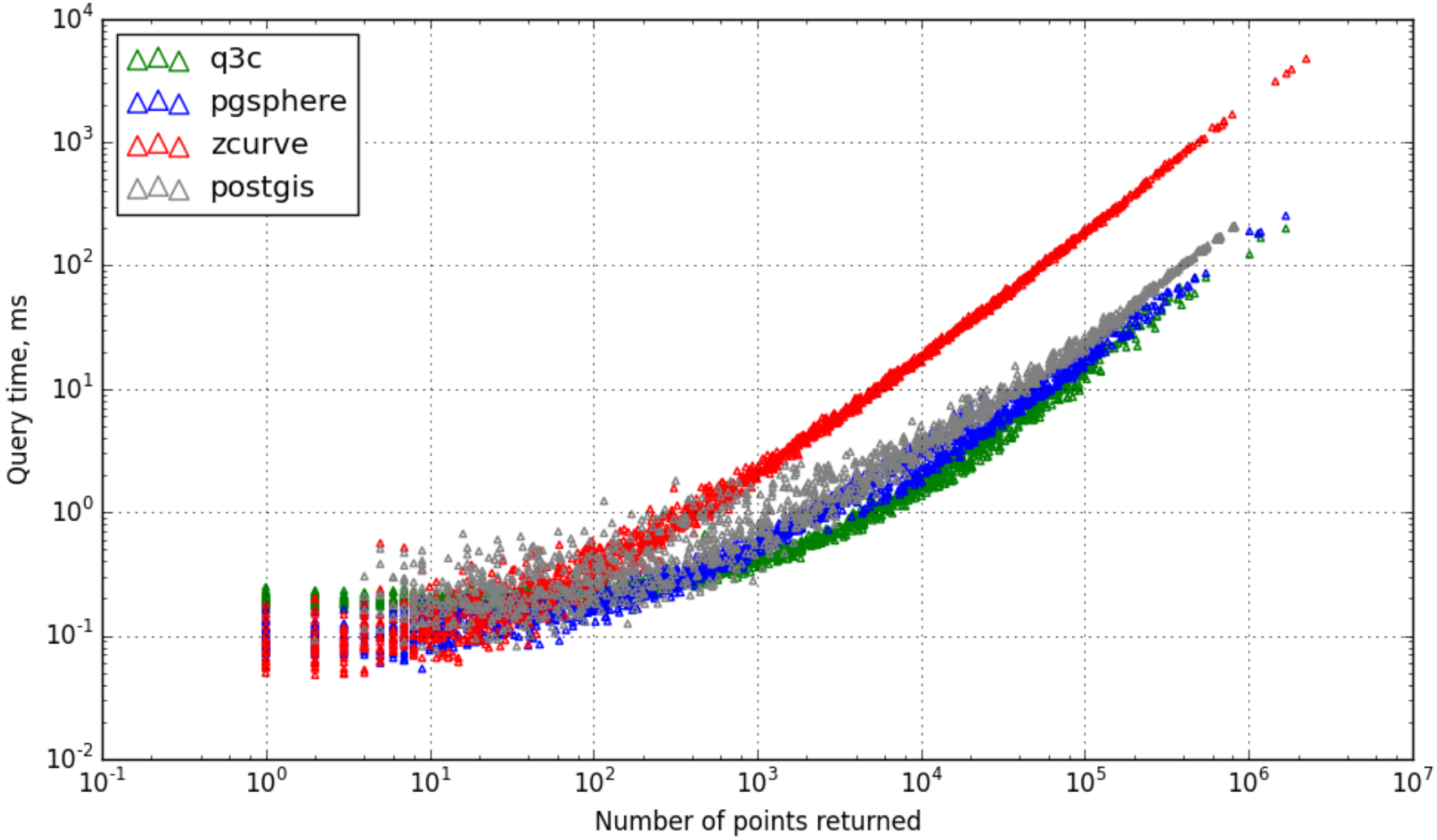
Index size, GB



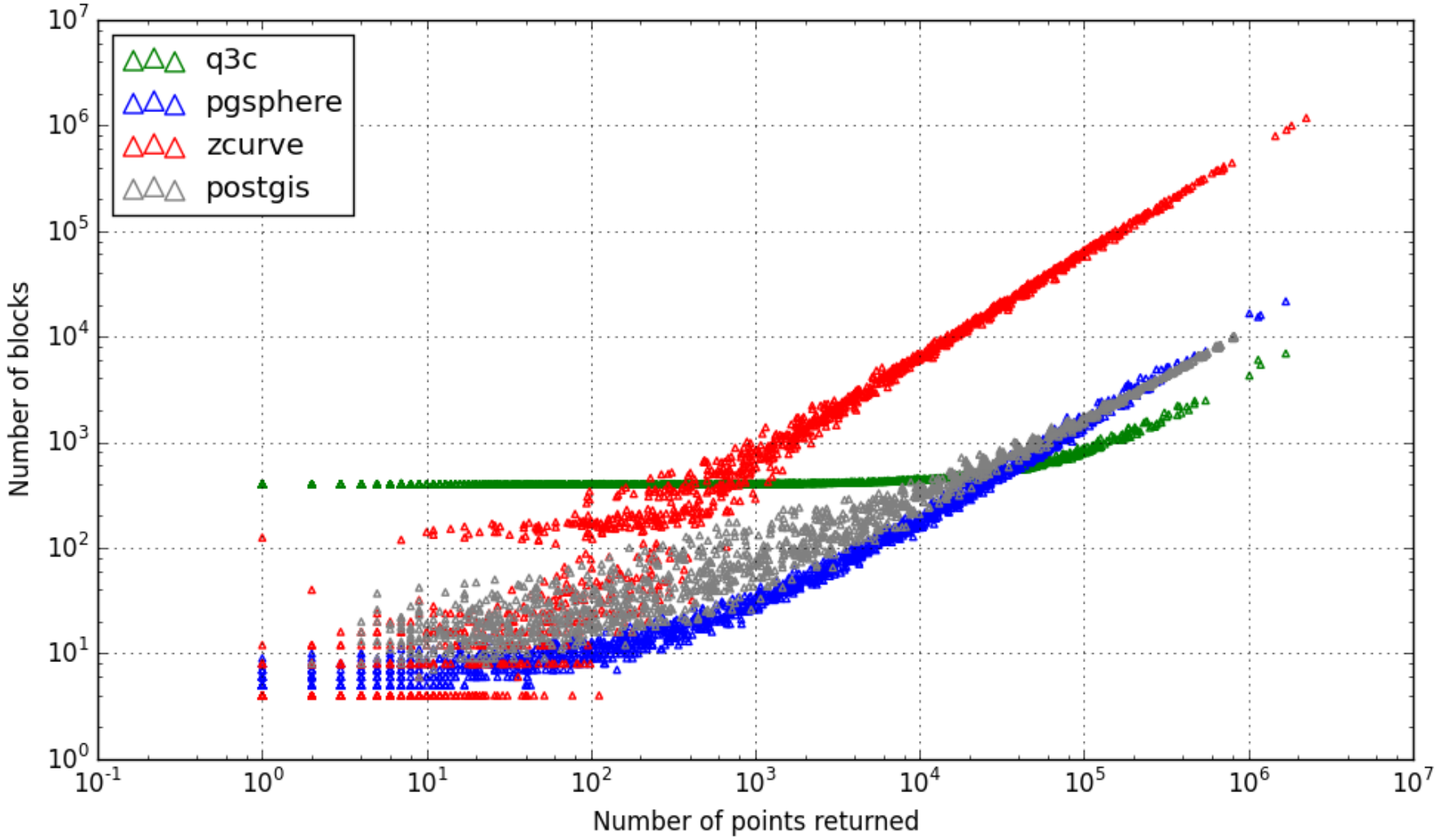
Build time, sec



Radial query time



Radial query blocks used



Conclusion

СУБД PostgreSQL provides rich set of features for working with spatial data including spatial objects in spherical system of coordinates. It allows to effectively process various types of search among them.

Future work

- Improve z-curve
- Index using GiST and z-curve in spherical coordinates
- Implement HTM using SP-GiST

Postgres PROFESSIONAL

Thank you for attention!
Any questions?