

The Power of PostgreSQL and PostGIS - Interview with Paul Ramsey



In this article, we speak to Paul Ramsey about how both PostgreSQL and PostGIS can provide a solid foundation for any small to large organisational GIS architecture. Paul, a Solutions Engineer at Carto, has been working with geospatial software for over 15 years: consulting to government and industry; building a geospatial software company; and programming on open source software.

He co-founded the PostGIS spatial database project in 2001 and is currently an active developer and member of the project steering committee.

In 2008, Paul received the Sol Katz Award for achievement in open source geospatial software, and he speaks and teaches regularly at conferences around the world. In the following discussion, Paul provides some insights into the market position of PostgreSQL and PostGIS compared to that of big-name vendor options.

GIS Professional: If a GIS manager is planning a from-scratch installation, the first concern should be the GIS storage layer - is this a fair assertion? Or is there something else you think should take place of preference as the GIS Stack foundation?

Paul Ramsey: Building technology infrastructure around a database is a good practice, but only if the staff feels comfortable and competent with it. Otherwise, it's a magical box in the corner and requests start getting bottlenecked with a small core of staff technology wizards. If everyone doesn't feel some ownership of it, feel that they get value from it, it will just breed resentment.

So folks have to see the advantage in terms of building multiple channels to data (web, desktop, analysis), in terms of data integrity under concurrent use, and in terms of in-place analytical power (spatial SQL queries). For a "GIS shop", that means that staff who are accustomed to reasoning about data visually or via a GUI will probably need some training in SQL, preferably around use cases that demonstrate the advantages of breaking away from the desktop.

That said, the architectural advantages of a database-centric set-up are pretty obvious: uniform access to data via a standard query language, one canonical working copy, concurrent access, easy automation of common tasks, and standards-based connections between systems. Without staff buy-in though, all that is for naught.

GIS Pro: Files or databases? In the old days, we said 'store your vector data in a database, your raster data in the file system'. Does this mantra still hold true? Are there any scenarios that come to mind that might make storing all your data as files, or storing all your data (including rasters) in the database?

PR: Just talking about vectors, if my team was small enough, using a decent file format like GeoPackage for all the data could make sense. The danger is really one of versioning and process locking: one file means one editor at a time. That can lead to either multiple files or lots of waiting around, so an "all files" installation really has issues of scale.

For rasters, I don't think there is **ever** a situation where putting **all** your rasters in a database would make sense. For some select data, like elevation models or continuous modelling results, data that you might want to combine and analyse with vectors on the fly, having them co-located in the database with the vectors can unlock interesting capabilities. But **all** rasters implies having visual data in there too, stuff that's really only used for backdrop and eyeballing, and that's just silly.

GIS Pro: What do you think are the key characteristics of an 'enterprise ready' database? I mean this particularly from the point of view of someone making a purchasing decision between some of the other major players - Oracle Spatial, Microsoft SQL Server, MySQL, MariaDB.

PR: There are a few things that distinguish an "enterprise" from the average guy on the street: the quantity of data, the number of

simultaneous users, the need for security and access control, and the willingness to pay extra for the word "enterprise".

At this point, I feel that the only thing PostgreSQL is missing relative to the other leading brands is a face-to-face sales force out knocking on doors and taking people to lunch. All the technical aspects -- scale, concurrency, integration with authentication/authorisation directories, cryptographic support, granular security roles and row-level security -- that's all there. If there's concern about "enterprise" PostgreSQL it's almost certainly a perception issue, not a features issue.

GIS Pro: Can you speak about the cost of ownership for PostGIS - are there any formal studies done which illustrate the comparative cost between different spatial data stores? Or are there any useful anecdotal case studies out there?

PR: The software cost for PostGIS is certainly lower, starting from the zero-dollar acquisition cost. Once you get into running it, the external support cost can be as low as zero for organisations that feel safe without a net, into the thousands for commercial PostgreSQL support from companies like EnterpriseDB or 2ndQuadrant. They are almost always a fraction of equivalent big vendor maintenance prices though.

The staff cost of ownership is a weird one. On the one hand, the base cost of someone who describes themselves as a "Proprietary DBA" is always going to be fairly pricey. There's also an assumption that a DBA does "DBA things" exclusively. Generally, people don't DBA a 'proprietary' database instance off the side of their desk. So there's an "organisation size" filter right away in the market for the staff of these proprietary companies.

The flip side is that, while lots of people have enough PostgreSQL experience to keep a modest installation up and running and backed up, the population of people with deep PostgreSQL administrative DBA experience is still small and growing. At this point, the best bet for getting a full-time "PostgreSQL DBA" is actually to hire a DBA of a proprietary company and have them train themselves up on PostgreSQL details, as the broader DBA concepts transfer over just fine.

GIS Pro: A GIS manager planning a deployment, trying to choose an appropriate geospatial datastore, and looking at PostgreSQL/PostGIS as an option might take some comfort from knowing that it has been used in large deployments by other organisations elsewhere. Can you give some examples of where this is the case? And what kind of data volumes have been managed in these settings?

PR: The example I keep coming back to is one of the oldest: the French National Mapping Agency started managing their national planimetric base in PostGIS over 10 years ago. At this point, it's practically a legacy system. That's a big operational data set, over 150 million features, and I think they chose PostGIS in a fair comparison over the likes of proprietary tools.

In terms of size, there are a number of organisations in the UK that load and work with the Ordnance Survey MasterMap product in PostGIS. That's a data set of over 500 million features. Most GIS managers don't have a problem of that scale to deal with.

And of course, it's worth mentioning that OpenStreetMap runs their service on a couple of beefy PostgreSQL servers. That's a world's worth of data under continuous editing load. They don't use PostGIS spatial types for their raw data model, but PostGIS is used behind the rendering chain that spits out the visual representation of OSM you see on the web.

GIS Pro: Finally, can you speak a bit about the future and where the demand may be, particularly in terms of addressing the unstructured data trend?

PR: NoSQL turned out to be a bit of a flash-in-the-pan, and PostgreSQL learned the two important lessons it needed from that moment: first, treat unstructured data as a first-class citizen, with a JSON data type; and second, be prepared to scale horizontally. Thankfully, developments in terms of JSON support for building non-schema constrained apps that are not schema-constrained, the development of a scalable multi-master PostgreSQL extension, and improved support for partitioning and foreign data access in the community development have changed all of this.

This article was published in GIS Professional April 2018