

GeoSpock: bringing data up to warp speed!

Relational, hierarchical, network and object-oriented databases have all played their part in data processing and analysis. More recently, a new breed of spatial and temporal indexing is being exploited to keep pace with our everything-connected world. GeoConnexion spoke to GeoSpock CEO, Richard Baker, to find out more

GeoConnexion (Geo): It was GeoSpock founder Dr. Steve Marsh's 2013 breakthrough in simulating neural networks that paved the way for this new approach to data manipulation. Can you explain how this has been exploited?

Richard Baker (RB): Steve's pioneering work in simulating human brain function in real-time was based on his creation of HiveMind, a super-compiler for the scalable custom BlueHive supercomputer hosted by Cambridge University's Computer Architecture Group. The efficient programming model stemming from this research now powers GeoSpock's cloud-based tools. These can analyse trillions of geospatial and temporal data points and extract and present meaningful answers in sub-second response times. For many, it represents the go-to solution for processing the petabyte-scale data volumes associated with smart cities and the Internet of Everything (IoE), including autonomous vehicle fleets. As well as location-tagged data, the toolset also has application in weather modelling, facial recognition, DNA sequencing, and many other extreme data-intensive tasks.

Geo: You mentioned GeoSpock's cloud-based tools. Can you say what these are and what they do?

RB: We currently have three SaaS tools on offer: infin8: a geo-temporal data indexing

solution that processes and organises data for sub-second data retrieval. illumin8: a real-time, interactive data visualisation and analytics tool
extrapol8: providing general and specialist spatial analytics capabilities against a GeoSpock dataset stored in infin8

Geo: Presumably the service offering is based on providing a high level of support?

RB: In many cases, clients don't just want to buy the software to host in their own cloud environments. They are really looking for professional services, data science consulting, help on interpretation, and so on. To assist them in this, we've developed City as a Platform for smart city applications and Location Intelligence as a Service for Enterprise customers. In the Enterprise sector, we've identified three key markets – automotive, retail and telecommunications – where location intelligence and location metadata is critical to business success.

Geo: Within automotive, the advent of autonomous vehicles poses huge data processing challenges. Does GeoSpock have a role to play here?

RB: Simultaneously managing a million autonomous vehicles driving around a city of the future is essentially a HiveMind



Richard Baker Photo: GeoConnexion

problem, i.e., where each vehicle needs to understand its contextual and spatial relationship to the city and to every other vehicle. As such, we're working with OEM auto manufacturers in the UK, US and Japan to tackle challenges of this nature.

Geo: Can you say something about pricing?

RB: In terms of smart cities – and there are now 18 such projects running in the UK – we've priced our offering at £60,000 a year for a typical one terabyte starter package. Available via the Government's G-Cloud procurement gateway, this includes all of our analytics, programmatic and data management services and is designed to help customers put their legacy data into one operating system environment as a prelude to bolting-on their next-generation infrastructure. Without the holistic view of temporal and

geospatial data enabled by this approach, you really lack the intelligence and analytical understanding that is needed to address smart city challenges and opportunities.

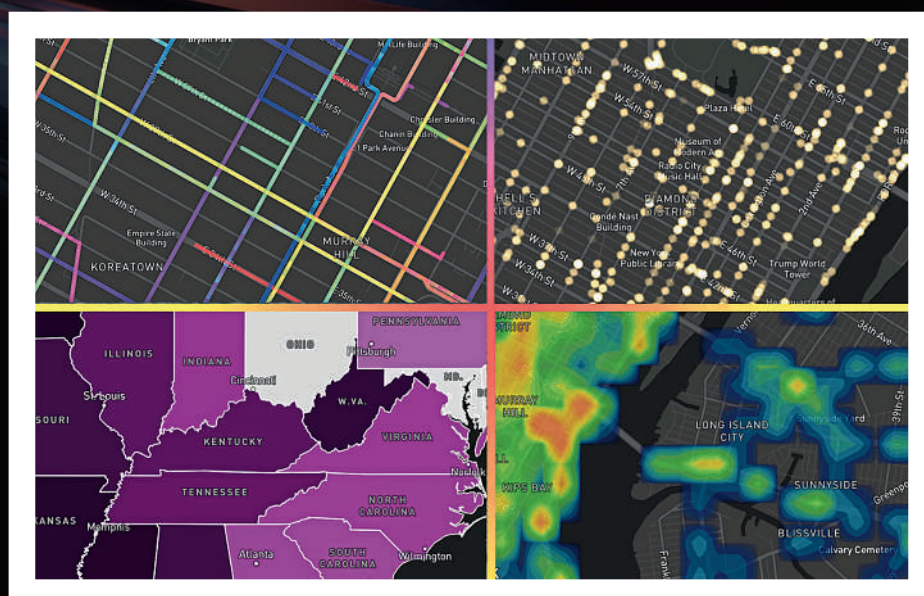
Geo: You've spent the past two years getting the business up-and-running. So how has it evolved over this period?

RB: At the end of 2017 we had a 13-strong data science and software engineering team in place and, in early 2018, established additional teams to handle operations, sales and marketing and product management, bringing our total workforce up to the 45 mark. We launched our suite of products and solutions in the Spring of last year; opened our London office in June, and extended our operations in mainland Europe and the Far East.

Geo: The company has also been busy forging business partnerships, most recently with GeoWorks, an industry centre established by the Singapore Land Authority (SLA). Can you say something about this?

RB: We will be working with GeoWorks in its drive to create a vibrant geospatial ecosystem for Singapore ... one that will promote business growth and innovation. In doing this, we will utilise the SLA's OneMap datasets for geo-temporal visualisation on GeoSpock's platform to accelerate business intelligence and insights. The collaboration will also explore the use of our technology as a 'data ocean' for multiple disparate geo-temporal datasets, as well as its use as a fast, lightweight 3D data processing solution, not least for an envisaged Smart City Operating System (OS).

Geo: There is much talk of GPU in-memory databases being a more capable and cost-effective alternative to CPUs for the next generation of database and big data analytics



Examples of geo-based visualisations from GeoSpock-processed data points include (from top left clockwise): Tracks, Point Density, Heatmap and Choropleth. Images: GeoSpock

applications. What are your thoughts on this?

RB: It's true that cloud environments housing GPU arrays are being positioned for such purposes. We don't agree with this approach for the simple reason that current GPUs are limited in terms of memory, with the biggest holding only about five billion rows of records – not a significant figure by today's standards. Beyond this, you end up having to virtualise the next GPU in an array of database instances. With such arrays for geospatial applications typically containing around 20 GPUs, this roughly determines the upper limit. So we're a long way from dealing with petabyte - let alone exabyte - volumes of data. For us, it's rows under management that poses the bigger challenge in simultaneously indexing temporal and geospatial data.

Geo: So what is your solution to this?

RB: By paying great attention to how we apply our indexing methodology in spinning disk environments we can extract much greater performance out of commodity-grade hardware. The result is sub-second response times on datasets containing trillions of rows of records and petabytes of data.

Geo: As we went to press, we received news that GeoSpock had secured £10 million in a further Series A+ funding round, bringing the total raised to date to £19.25 million. How will this latest investment be used?

RB: It will be used to support our rapid international expansion strategy, particularly in the strategic markets of Singapore and Japan. The investment will also allow us to continue investing in research and development, particularly in the areas of machine learning and data science.