IT for CEOs & CFOs is published by House of Words Media Limited.

The current issue and full text archive of this journal is available on our main website at **www.creditcontrol.co.uk**.

The Positive Impact Of Graph Technology

Emil Eifrem

Emil Eifrem
CEO
Neo4j

Biography

Emil Eifrem is CEO and co-founder of Neo4j (http://neo4j.com/). Previously Chief Technology Officer of Sweden's Windh AB, where he headed up the development of highly complex information architectures for Enterprise Content Management Systems, Emil famously sketched out what today is known as the property graph model on a flight to Mumbai in 2000.

Emil is a frequent conference speaker and a well-known author and blogger on NoSQL and graph databases, as well as co-author of the agreed Bible on graph databases, O'Reilly's Graph Databases (http://graphdatabases.com/).

Keywords Graph, Graph databases, Big Data, Connected data
Paper type Research

Abstract

Graph technology has come a long way. In this article the author considers what propels this success and the positive impact this technology is making.

Introduction

Graph technology has come a long way – from powering investigative journalism in the Panama and Paradise papers to helping in the search for a cure to cancer, even helping speed up our exploration of deep space.

What propels this is its success with helping developers better understand data relationships. It's quintessentially a technology optimized for finding and tracking connections – and in a context where the volume of data keeps exploding, so is the value of the connections businesses want to extract from it.

So last year's Paradise Papers[1] revelations and their 2015 predecessors, The Panama Papers, have put under the spotlight a complex web of the 1% financial dealings, achieved through probing, sifting and managing huge volumes of data. At 2.6 terabytes and 11.5 million documents, the Panama Papers alone was far larger than anything Snowden or Wikileaks ever managed, while its 2017 successor was not far behind, at 1.4 TB of data and 13.4 million documents.

These exposures of the activities of clients of offshore law firms qualify as the world's largest financial investigative journalism probes ever. Why was this technology so suitable for this landmark work? Graph databases excel at spotting

data relationships at scale. Consider the case of the former Icelandic Prime Minister, Sigmundur Gunnlaugsson, who was forced to step down after revelations that his family hid millions in offshore accounts. Gunnlaugsson was connected to the offshore accounts only indirectly, several steps removed, via his wife.

That attempt at hiding his trail using layers of connections is highly typical of such cases, and such connections remain invisible using a standard relational database approach.

Meanwhile, one of the latest investigative journalism projects that graph technology is powering is research into Russian Twitter trolls. US broadcaster NBC ran a recent story based on mining of a set of 200,000 deleted Tweets sent out by suspect sources into how this form of fake news might have affected the 2016 US Presidential election[2]. It found that the key to detecting fake news is spotting those hidden connections between accounts, posts, flags and websites.

Also note that a highly significant fact to these data investigations is that they are carried out not by University PhDs, a big consultancy or an IT team at a major software firm, but a tiny team of data-trained reporters.

Graph and its role in medicine

That's the first story of the impact of graph technology. The second story is about its impact in the field of medical research. It turns out finding a cure for cancer is a very multi-faceted problem with a lot of moving parts: factors include things like what compounds you use for your medical research, what segment of the population you are targeting, their demographic, their genetic history, their medical history, their lifestyle, whether they are smokers or non-smokers, amongst other aspects.

All these things affect and influence each other. What you are looking at is a massively connected data problem of factors such that if you change this one component, then through a butterfly effect it will cascade across the network. The challenge is that medical research institutions have data stored in disparate silos – or even in different institutions, so data on genetic makeup is in one database, and data on diets and lifestyles in another database, and you can't easily cross relate these data sets with traditional tools. The problem is compounded by the fact that these institutions are often wedded to relational database technology, which is extremely poor at analyzing connections.

By contrast, the USP of graph database technology lies in discovering relationships between data points and understanding them – and at huge scale. That is why it is ideal at allowing the medical researcher to uncover hidden patterns when they are looking at the hard problems out there, and may thus prove critical in powering cross disciplinary research and real medical breakthroughs.

The third example of the huge impact graph technology is having concerns NASA's Project Orion deep space exploration vehicle[3], based in the Johnson Space Centre in Houston, Texas. A couple of years ago the Orion project ran in to a significant

IT for CEOs & CFOs is published by House of Words Media Limited.

The current issue and full text archive of this journal is available on our main website at **www.creditcontrol.co.uk**.

problem; in all of its test simulations, the return capsule ended up being upside down or sideways on.

It turns out that there is a part on the return capsule called the upright construct, which ensures it is facing in the right direction, and the operating construct here was faulty. This is not NASA's first space mission, however – and over the last 50 years it has built up a lot of corporate knowledge about how to do this. As a disciplined engineering organization it has a Lessons Learned Database containing evidence built up over a long time of all the things used in engineering, both the things that have succeeded and the things that didn't.

It turns out that Project Orion's upright construct shares the same DNA as the 1960s Project Apollo, so maybe the solution to the problem lies in the files relating to Apollo? The challenge was that the dataset was too huge to easily navigate to find out this insight. After considerable time was spent trying to glean something from the database, which proved fruitless, the chief knowledge architect decided to move the data into a knowledge graph. In just three hours, he had found 20 potential documents that could be related to the problem – and soon, NASA discovered the solution.

The mission was back on track, just another example of how graph technology is having a positive impact on our world.

Another huge impact graph technology is having on our lives is in the area of Machine Learning (ML) and AI. A lot of AI and ML is actually based on graph, and a lot of the algorithms that underpin it can be expressed as graph data structures. Also, while Google's dominant search engine has always been driven by smart software, it has recently become even smarter after incorporated AI and expressing its underlying architecture in a Knowledge Graph.

In conclusion

Google isn't the only company using the power of Knowledge Graph as a way to complement its underlying software – look at the now almost ubiquitous shopping or customer service 'bot'. eBay's AI-powered ShopBot, for example, is built on just such a graph representation.

Graph is having a huge impact, smart developers agree. Is it time to see what positive impact this technology could make on your project or organization?

Reference

1 https://www.icij.org/investigations/paradise-papers/
2 https://www.nbcnews.com/tech/social-media/now-available-more-200-000-deleted-russian-troll-tweets-n844731
3 https://www.nasa.gov/exploration/systems/orion/index.html