**Emil Eifrem** explains how database technology that is being adopted by the corporate world can help the Government make sense of its data to help create a safer society

# GRAPH DATABASE HOLDS KEY TO GOVERNMENT SECURITY

Big data is a big opportunity for the Government, but it does not come without its security challenges as its volume grows at an exponential rate. Recently a new database, dubbed NoSQL, has started to garner attention from CTOs and CIOs looking for an answer to safety issues.

You may have already come across NoSQL, also referred to as 'Not Only SQL', in the context of big data and its skill in organising and analysing high volumes of disparate data types in real-time (and sometimes referred to as 'big data technology', in fact). But its expertise does not end there.


©Getty Images

***Graph database technology has the power to quickly work through duplicate names and IDs to support border control officers***


According to IBM, we produce around 2.5 quintillion bytes of data on a daily basis – with 90 percent of the world's data mountain having been created in the last two years alone. This seemingly endless stream of information is collected from everywhere, from the internet to social media and sensors. Market analysts IDC estimates that the world's data will hit 44 zettabytes by 2020, with Internet of Things (IoT) accounting for 10 percent of the total. The volume of data will increasingly outpace storage facilities. But how will it be used and can it really help Governments tackle major issues such as national security and crime?

As venture capitalist Mary Meeker pointed out in her annual *Internet Trends* report this year, we are collecting more data than ever before and this has made data a serious growth platform. At the same time, the public has become increasingly worried about where its data is and how it is being used. Governments want to hone as much useful information as they can from big data, while protecting the privacy of individuals.

It is only natural while accumulating large amounts of data, to want to know what is in it – and the Government is not any different. Much of this data is semi- or completely unstructured, which makes it difficult to decipher. This is a major issue for law enforcement agencies and security experts, as they need a way to unravel big data. If we had a better way of interrogating data sets, isolating connections and joining the dots between relationships, we would be ahead of the curve in crime prevention.

The biggest hurdle that police, security agencies and border controls have with relational database management systems (RDBMS) is that they simply can't read data stored in unstructured formats such as that siphoned from social media. In addition they haven't got the ability to make implicit connections within the data.

So it is little surprise there has been a surge in interest in post-relational databases that can unmask big data. Each type offers particular strengths, but what underscores them all is their dexterity in processing large volumes of data, in particular semi-structured or unstructured data.

Graph databases provide an excellent way of storing data that refer to things that are naturally connected in the real world, such as people in social network groups or vehicles licensed in a particular year. Storing individual data pieces as 'nodes' in a graph and linking them enables the Government to quickly get new views on data that would be almost impossible with relational data structures such as neighborhood demographics, number of people living in a region with foreign bank accounts *etc.*, making it an ideal solution for cybersecurity and fraud detection.

Early this year, for example, we saw the International Consortium of Investigative Journalists (ICIJ) reveal the results of research into supposed illegal offshore transactions that were allegedly made by Panamanian law firm Mossack Fonseca on behalf of wealthy and highly influential business people. These hugely complicated transactions were unraveled using graph technology.

Of course, RDBMS technology still has a very key role to play in Government IT, as do earlier database formats such as flat-file/network and CICS. But in an increasingly digital society, with much centered around digital identities, a far deeper understanding of big data is necessary.

To do this, just as in the enterprise world, we need to be able to connect the dots to map what traditional enterprise Business Intelligence (BI) characterise as the '360-degree' view of the digital customer or digital citizen. Graph database technology can provide Governments with a much more detailed picture of its citizens and society as a whole. Data gleaned by border controls for example, can be used to better address human trafficking and terrorism threats, for example, by linking back-end data sources.

While NoSQL may be able to handle data at scale better than RDBMS, graph databases in particular can work with the connections between people, places and events that provide the foundation of what happens in the real world. Using traditional relational database technologies requires modelling these connections as a set of tables and columns, then making a number of complex joins and self-joins. These queries are often very difficult to build, expensive and slow to run, while scaling them in a way that supports real-time access creates some major technical challenges. But the connected analysis that graph databases supports is the main area of interest here, as today's digitally empowered public services depend on being able to spot these connections for legal, security or improved service delivery, for example.

There is a number of ways graph databases are contributing to security. Take border control, for example. How do border control officers know that a passport is official or supporting a fake identity?

One country we work with has a high rate of immigration, but does not automatically give working or residence rights to all who cross its borders. It often gets people representing themselves again and again, using slightly different names on their applications, in a bid to gain access. Border control needs to sift these rejected applicants, who are trying to dupe officers, from those who have made a first-time, legitimate application.

Graph database technology has the power to quickly and thoroughly work through the enormous number of duplicate names and IDs in real time, which supports enhanced decision-making for border control officers.

Another project we are working with involves using telecom data to better protect us all. Telecom data can offer useful information for police and security experts because of the connections graph databases are capable of uncovering. For example, if a group of

# GRAPH DATABASE HOLDS KEY TO GOVERNMENT SECURITY

©Getty Images

***Market analysts estimate that the world's data will hit 44 zettabytes by 2020, making data centres increasingly important***

people is continually calling the same foreign number and consistently, it may indicate some sinister activities are afoot. Alternatively, internet data may spot young people that are being radicalised through social media. If you add in criminal data records, further dots can be connected, which may provide valuable information for international Governments to work together to protect security.

Government cybersecurity teams have huge amounts of data to analyse, ranging from network logs to email communications, which are generated by systems monitoring data. The problem again is that it is unstructured and flows in from heterogeneous, incomplete data sources. Because of its complexity, it is virtually impossible for tabular solutions to analyse. Graph databases, however, make it easy to store and query the data, regardless of its size. It can for example, be used to spot patterns that may lead to Phishing scams where a criminal pretends to be a trustworthy organisation, such as a bank, in order to gain access to sensitive information.

Government today collects and distributes an enormous amount of data. Its increasing complexity and sources makes it difficult to get a cohesive picture. Metadata, however, is capable of providing a 360-degree view. This is why graph databases are being evaluated to deliver national-level Master Data Management (MDM) to connect big stores of disjointed sets of information and present them in a unified way. Master data is mission-critical, operational data about citizens, services and organisations, for example, including the relationships between them.

Graph databases can also help build a registry-style MDM that stores the most useful metadata including the location of the actual master data. With a single trusted view on all its data, Government agencies can increase efficiency and address increasing security threats.

The implications for better protection against terrorism and organised crime, for example, is enormous. Unnecessary questioning or controls can be avoided for those who are abiding by the law.

Based on conversations with the public sector we have had, we know that these innovative projects are just scraping the surface of the capabilities graph database and NoSQL have to enhance the way the Government works to ensure the safety of its citizens.

We are talking about a system so well informed about a Government's citizens, their behaviour and mobile phone use that it can, in seconds, track back to anyone who dials in a bomb warning, a kidnap call or an illegal immigrant meeting point for a shipment time. Graph databases can provide interactive maps of cities, detailing which crimes are being committed in which neighborhoods, so that police can use fact-based strategies to allocate resources or track bank accounts that may be linked to crime rings or terrorism. The scenarios are seemingly endless.

Graph databases are impossible to beat on certain types of big data, such as that gleaned from social networks, cellphones and IoT devices, making it a remarkably powerful tool to help the Government make our world a safer place to live, work and travel.

**Emil Eifrem** is co-founder and CEO of Neo Technology, the company behind the world's leading graph database