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**.

Legislation and Compliance

Leveraging the Power of Graph Technology for GDPR Compliance

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 database, Graphs, GDPR
Paper type Research

Abstract

A seismic shift occurs in the data management world in May 2018 when the General Data Protection Regulation becomes law in the European Union. The new GDPR regulations that take effect on 25 May 2018 will apply to all EU and foreign organizations handling personal data of EU residents. They mandate strict compliance and call for steep fines for privacy violations. If you commit infractions or are subjected to random checks, regulators will require you to prove your compliance with GDPR requirements. In this article, the author explains why he believes that graph database technology is the best way forward in reaching General Data Protection Regulation (GDPR) compliance.

Introduction

Countdown to GDPR has started and there are only months to go before the regulation replaces the Data Protection Directive of 1998. The General Data Protection Regulation (GDPR) will apply from 25 May 2018. The government has already confirmed that the UK's decision to leave the EU will not affect the start date for GDPR. Its arrival, however, will bring sweeping changes for businesses dealing with personal data.

The worrying aspect of GDPR, however, is how few CIOs are actually prepared for the regulation, which has sharp teeth and brings with it hefty fines for those that don't comply. A survey[1] in June this year by consultancy PwC, for example,

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**.

highlighted the fact that of 150 GDPR readiness assessments it carried out with clients, many found it difficult to explain their preparations. Many respondents did not know how to move programmes beyond risk reviews and data analysis, leaving them open to data protection fines unless they could enact radical and fast changes to their business plans.

All is not yet lost, however. There are ways of utilizing data today that could reduce the time it takes to get GDPR compliant and relieve a few serious headaches for CIOs. To be fully GDPR-compliant, companies must be able to track the journey of a customer's personal data. 'Personal data' means any information relating to an identified or identifiable person referred to as a data subject. This now covers online identifiers such as IP addresses and mobile device IDs. Companies must record exactly when any data was acquired, whether consent was given when it was obtained, where and when it is moved and on which systems it resides. They must also explain how exactly this data is being used.

Figure 1: A modern graph approach is far superior for GDPR

GDPR Task	Traditional Approaches	Modern Neo4j Approach
Trace data through enterprise systems	Complex queries with hundreds of join tables	Simple single query traverses all enterprise systems
Preserve the integrity of data lineage	Broken data paths and lineage, especially with NoSQL databases	Continuous, unbroken data paths at all times
Effort required to add new data and systems	Days to weeks to rewrite schema and queries	Minutes to draw new data connections
Time to deployment	Months to years	Weeks to months
Response time to GDPR requests	Minutes to hours per query	Milliseconds per query
Form of GDPR responses	Text reports that are not visual and prove very little	Visuals of personal data and the path it follows through your systems
Bottom line	Long, ineffective and expensive	Easy, fast and affordable

Appropriate technical and organizational measures must be taken to secure this data. If there is a data breach, the company must be able to describe its nature, steps taken to control it and possible consequences.

Data logistics create an enormous challenge. In our digital world data is spread across multiple locations, both on in-house networks and with cloud service providers. GDPR holds companies accountable for all the data, regardless of where it is stored. At the same time, it requires that all organizations can access, report and remove personal data from all systems when called upon to do so either by consumers or regulators.

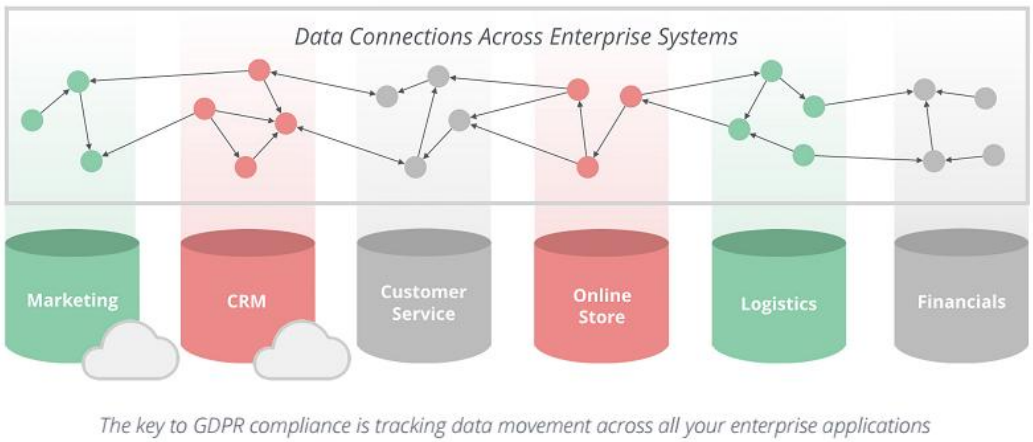
IT for CEOs & CFOs is published by House of Words Media Limited.

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

Connections are pivotal to personal data

Connections are fundamental to data. Organizations need to be able to pull up entire vistas on the data they hold on customers. This might sound like a burdensome task, but it can be done quickly and efficiently by tracking the connections between the data.

Figure 2: Tracking personal data movement



Attempting to track these connections with SQL isn't feasible as the data model is too restrictive for the demands of GDPR. Personal data usually follows an unpredictable road through an organization's infrastructure. This winding route, however, can be captured using graph database technology. Neo Technology's latest whitepaper[2] *The Fastest Path To GDPR Compliance* explains this in more detail.

The power of graph databases has long been recognized by the web giants, Google, Facebook and LinkedIn, for example, and more recently was harnessed by investigative journalists sifting through gargantuan amounts of data to uncover the Panama Papers scoop. Graph databases by their very nature are suited to connected-data applications such as those that underpin GDPR, where data relationships have just as much value as the data itself.

So how does it work?

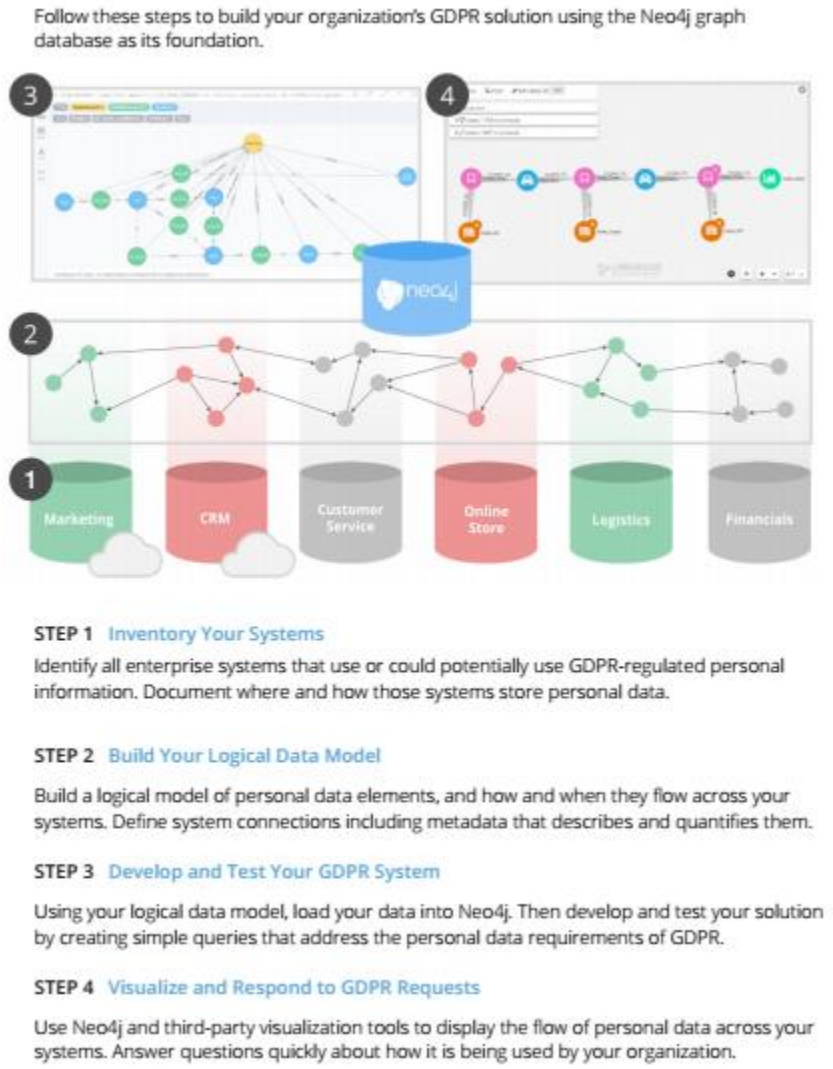
A traditional approach to tracing data through enterprise systems requires complex queries with literally hundreds of join tables, which often also result in broken paths. Using a graph database, however, a single query can cross all of an organization's systems, providing you with a continuous chain of data paths.

Of course, traditional relational database technology has its place. It is good at handling highly structured datasets that change infrequently and have minimal numbers of clear connections. But this isn't a suitable solution for GDPR, which demands intricate tracking of personal data.

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**.

Figure 3: Four steps to GDPR compliance



GDPR project leaders who are looking at real-world tools that can help them solve the GDPR conundrum would be advised to take a close look at graph database technology. We aren't just talking large global organizations here either. But beware database vendors touting graph capabilities. These are hybrid databases that come with trade-offs and often their own non-native issues.

If you are looking at getting GDPR compliant, graph technology is definitely one of the tools you should examine. It will take the stress out of managing personal data, be it on a regional, countrywide or global level.

Reference

1 http://www.consultancy.uk/news/13562/uk-fines-for-data-breaches-double-to-32-million-gdpr-to-lift-bill-further
2 https://neo4j.com/resources/gdpr-compliance-white-paper/ ?ref=sarum