Rising to the data analytic challenges of new data regulations

General Data Protection Regulation

Recent developments in data protection law such as the General Data Protection Regulation (GDPR), and increased demands on creating responsive services, are combining to increase the burdens on data owners and processors, in a number of ways:

Increased data protection requirements

- Data owners will be liable for data loss even where the loss is due to the criminal actions of a third party, requiring greater **data security**.
- The **purpose** to which data may be put means that just because an organisation has a data item, it doesn't mean that everyone in the organisation can see it, nor that it can be used for any purpose.
- The **data life-cycle** is becoming more complex. Traditionally data would be created, used, and deleted (with varying degrees of success.) Now however, the data subject may wish to change, on the fly, the use to which it may be put. The organisation may wish to accommodate this, because it could give rise to further business. Or the data subject may wish to rescind permissions, and the organisation is obliged to comply.
- Data processing must be **correct** and shown to be accurate.
- Decisions made on data may need to be **explained** and justified to management, if not to a regulator.

which are enforced

These regulations are backed up not just with teeth, but the metaphorical equivalent of tactical nuclear weapons. Fines can be up to 4% of group turnover or 20 million Euro.

but still must deliver

The next generation of data analytic systems must be able to cope with these issues – whilst coping with ever greater **data volumes** and an enduring **need for speed**.

This is how a Scenario Analytics solution might address these challenges:

Increased data security

Considering first the requirements for **increased data security**; with the advent of GDPR it is no longer sufficient to hold your database on encrypted disks, because if an external hacker gains, or an insider has, access to the network, they have access to that data. Consequently, the database itself ought to be encrypted.

Ordinarily, that would mean to analyse the data in any way requires the data to be decrypted as it passes into the analysis system, and the results encrypted on the way out. While it is in the analytics system, the data has to be "in the clear" for the analytics to work. If the adversary gains access to the analysis system – or to the encryption keys it must be provided with to work – he has gained access to the data.

Homomorphic encryption

Scenario Analytics can indeed analyse the data without decrypting the data first, a technique known as homomorphic encryption. It splits the data into two components – a context without which the connections are practically unintelligible – and then analysing the connections independently of the context. The output is equally uninformative without the context.

Data purpose

The second problem revolves around the **purpose of data**. Data may only be used for the purpose for which it was collected, and for which permission has been given (which may change). Frequently, the purpose of use can be considered as a 'need to know' group, where that group is either a set of people, or a set of automated functions. This is particularly complex in an analytic environment since the results of the analysis may or may not be revealing. For example, the location of my bank transactions should be visible to the fraud department *if it changes significantly*, whereas the average value of my bank account (being as informative as my bank balance at any one point), should *not* be available to the fraud screening department (unless a potential fraud is discovered).

Data meta-data payload

Scenario Analytics carries with data a data meta-load, which can describe a number of data facets, potentially including who can see it. This can be used to decide which business function can see any particular result.

Complex data life cycles

The **data life cycle** – no longer consisting simply of data birth and death but now (metaphorically) including marriage and divorce – is very complex. This is especially so since it may change in flight.

Graphs and sub-graphs

In common with many tools, Scenario Analytics represents data relationships as a structure called a graph (items joined by links, such as cat-is-a->animal or xyz_co-makes->widgets). However, Scenario Analytics takes this a step further and the data meta-load may *itself* be a graph. This allows extremely complex data life-cycle and authorisations to be modelled. (An alternative representation, whereby life-cycle and authorisations are themselves complex items to which a data item links is equally feasible).

Data correctness

Next we have the issue of **data correctness** or, equivalently, uncertainty. Increasingly, tomorrow's systems will live in an environment of chaotic data, sourcing their data from a variety of systems with different degrees of accuracy.

Confidence levels

Scenario Analytics can keep the confidence level of any one data item as a data meta-load, and could then use it to calculate the degree of confidence to be placed on any one result. Indeed, in that way contradictions cannot just be tolerated but even exploited, for example in a fraud detection system.

Compliance

Systems must be able to **explain** why a conclusion was reached and Scenario Analytics can do this in natural language or with diagrams.

Performance

And finally **speed** and **size**. Scenario Analytics is able to store its data in a relational database (even when split for homomorphic encryption), and take advantage of the data management facilities these systems have developed over the years. It does not do this by converting its own 'scenarios' into SQL queries, an approach which has been shown to be inefficient, but by making a series of SQL queries as it builds up and refines its solutions. (Where appropriate, it will also build small transitory graphs in RAM.) As a result, it can be both big and fast, taking advantage of SQL indexes.

Primary Key Associates Limited

Secure Data Analytics

About us

Since 2010 Primary Key Associates has delivered world-class consultancy, developing and exploiting cutting-edge analytics, artificial intelligence and digital investigation technologies.

Answer-Focused Technologies

We re-invest much of our revenue to develop cutting-edge data analytics, investigation and artificial intelligence technology to address the business problems we see our clients facing. Our IPR portfolio includes:

- Primary Key **Scenario Analytics** – a fast, modern and flexible data analytics technology to find, explain and illustrate the 'unknown knowns' in your large datasets.
- Primary Key **Insight Engine** – Scenario Analytics combined with powerful entity and relationship extraction, applied to digital forensics that reveals previously unknown connections in evidence.
- Primary Key **Illuminate** – a social media analysis and intelligence technology and service to find and analyse open source data to address particular business problems.
- Primary Key **Incipients** – predictive analytics that identify what is going to trend.
- Primary Key **Distil** – a technology toolkit we deploy on client sites to both find frauds and identify the business practices that make you vulnerable to fraud.
- **Fast analytic heuristics** to overcome the bottleneck of solving intrinsically non-parallel, yet vital, problems.

Expert consultancy and services

Our team are experienced professionals in IT and related areas:

- As experts in enterprise and security architectures (including SABSA), information security and cryptography we help design, validate, and test systems which are secure both physically and in cyber-space.
- We build software systems from the smallest (in C and assembler coded microcontrollers) through the mobile (Java or Swift coded Android or IoS apps), to the ubiquitous (Python or C+ on PCs, web technologies and virtualised servers) and even the esoteric (real-time and spacecraft).
- We understand data (SQL and graph databases) and artificial intelligence (computational linguistics, image analysis and machine learning) and have fast algorithms for timely and secure analysis of big datasets and evidential analysis of social media.
- We provide cyber threat intelligence and competitive intelligence.
- We support civil and criminal digital investigations from open source research through digital forensic analysis and providing expert witness reports and testimony.
- We undertake both technical and business programme and project management.

e: feedback@primarykey.co.uk
w: www.primarykey.co.uk
t: +44 1403 599900

@pkaluk