Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

Pivotal™

Detecting Anomalous Behavior with the Business Data Lake

Data Science and Analytics helps you to move beyond the threats you can see to anticipate the threats that you can't

The most elusive security issues pose the biggest threats

Over the past 20 years, enterprises have been fairly successful in combating "the threats you can see". Measures have included antivirus, firewalls, access controls, and physical security of the data center, along with Governance, Risk and Compliance (GRC) methodologies. Some organizations have also introduced security operations centers that capture and analyze network traffic to discover threats.

What most organizations have yet to do is find a way tackle internal threats from employees and others who operate within their approved authority, but start to abuse delegated business privileges and rights. Staff may have been subjected to psychological manipulation, blackmail, bribery or be driven by simple greed to sell sensitive corporate information to nation states, competitors or organized crime. Examples such as Nortel Networks show the reality of this threat and its impact on the valuable IP of a company.

An example of the challenge is a system administrator who starts passing downloaded data to unauthorized individuals or organizations. Even advanced security solutions can't detect this, because the individual is accessing systems that they're allowed to access.

Organizations need a way to pre-empt attacks

Most businesses are already well aware that they need to identify anomalous behavior, behavior that steps outside the norm, before the damage is done, but until recently, this hasn't been practical.

The primary challenge has been that companies lack the analytic capability to identify anomalous behavior fast enough for action and reliably enough to be taken. This has been because historically the technologies and skills required to undertake what is a hugely complex mathematical task have been prohibitively expensive.

The second challenge around the information required for that analytics is not its existence but its availability. The information is fragmented across multiple systems used by different organizational functions and is often thrown away due the cost of storage and an inability to drive insight from it.

To address the challenge of anomalous behavior therefore two pieces must be addressed, firstly the mathematical analytics and data science must be available at a reasonable cost and secondly there must be a way which cost effectively consolidates the information required.

Capgemini's Anomalous Behavior Detection Solution

Capgemini now offers a solution that addresses these two challenges and can spot anomalous and suspicious behavior by applying advanced, machine learning algorithms to a broad and deep big data set. By applying innovative data science techniques to huge volumes of data, you can detect atypical patterns of behavior quickly, and often take action before a threat becomes critical. The solution:

- Learns what is normal, and the difference between what is "approved" (actions that someone should be taking in line with their role), what is "allowed" (actions that someone has the ability to take within their role, although not explicitly approved is allowed) and what is "suspicious" behavior that indicates a clearly different pattern of behavior.
- The Data Science model identifies anomalies and provides an initial score which allows you to categorize, prioritizing the risk for further investigation, taking action immediately or teaching the mathematical model that this behavior is either approved or allowed.
- Creates alerts so you can react before an anomaly becomes a problem.

The mathematical models allow the response to be tailored to specific behavior based on its risk to the business. Suppose an analyst building quantitative analysis models for a bank accesses a system that they have never looked at before. Because there could be a valid reason and model has learnt that the role regularly requires new information sets, it might be enough to notify a manager. At the other extreme, if a system administrator suddenly downloaded four terabytes of data, and the HR system simultaneously flagged up that they had just received a poor appraisal score, the system could switch off their access pending investigation.

A solution that addresses these two challenges and can spot anomalous and suspicious behavior by applying advanced, machine learning algorithms to a broad and deep big data set.

The solution can also detect events that are anomalous in IT, rather than human, terms. If someone who travels a lot logs in from Japan that might be fine, but if they had logged in from the UK only four hours before, then obviously the user may not be who they seem, and urgent action will be required to protect data and IP.

The advantage of using machine learning models is that they can be used to provide you with predictive analytics which alert you before any anomalous behavior occurs. The mathematical models become able to identify the trends that lead to fraud or a breach. This enables you to put in place new processes which can then react as a threat occurs to prevent it at source.

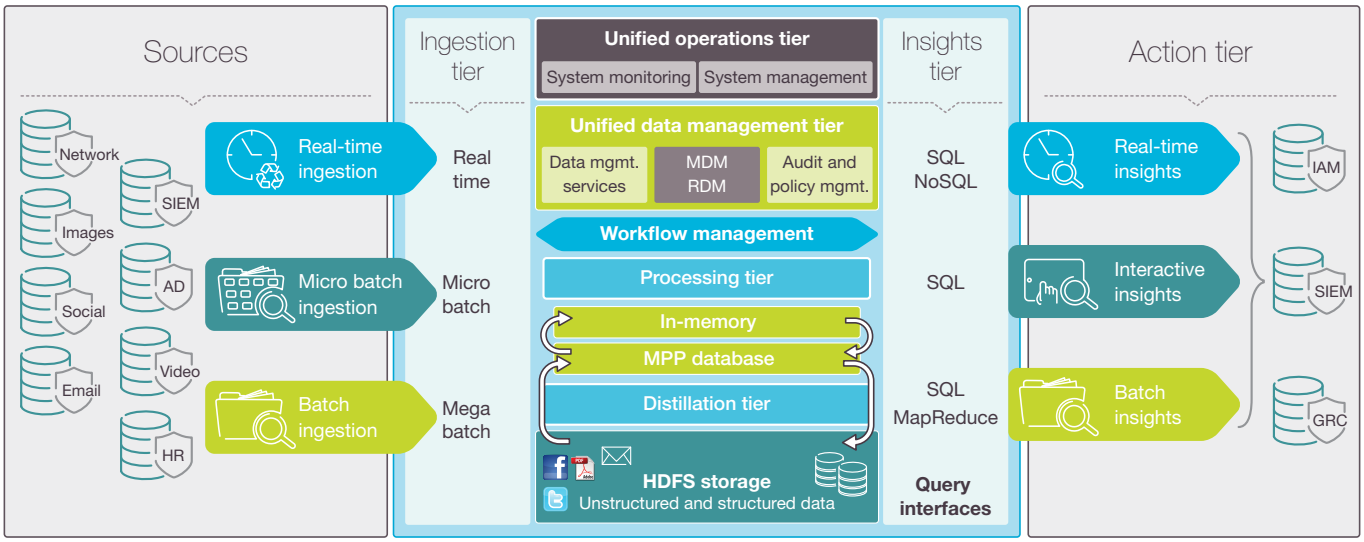
How it works

Data science, machine learning and the MADLib library are at the heart of this next generation analytics solution. The machine learning algorithms in MADLib are used to analyze the historical behavior of users and systems to understand what constitutes normal behavior; these algorithms are then weighted based on the risk of different threats to your business to create a baseline. By combining transaction, network, HR and even external data a view is created of both what represents the standard operating procedures for your business and what represents a business threat. The models can then be integrated into HR or security systems to enable you to react to threats as, or before, they occur.

All this depends on the ability to analyze huge amounts of data – both at speed and in batch modes. Capgemini's Business Data Lake (BDL) – a joint development with Pivotal – makes that possible. The BDL is able to complement existing Enterprise Data Warehouse (EDW) approaches, and in addition the BDL stores huge volumes of data, including unstructured data. There's no need for branded software or hardware – affordable commodity platforms can be used.

The BDL extracts and distils the views of the data that you need – whether these are "disposable" or shared across the business. It comes with analytic tools provided at a price point that wasn't previously possible. This means that data used for enterprise reporting can now be leveraged directly by rich data science applications.

Figure 1: Ingest structured and unstructured data to the Business Data Lake; create insight and take action



Benefits of the Anomalous Behavior Detection Solution

Protection from potentially catastrophic loss. This approach could save millions by preventing fraud – even more by protecting vital IP. It can counter IP leakage, theft of sensitive material, and abuse of company systems, and complement network-level detection by providing early warning of social engineering attacks. You can not only minimize exposure time and loss, but possibly pre-empt attacks. And you don't have to know what you're looking for – for one client, we not only identified a person known to be creating false records but also a number of others apparently using the same loophole.

A fast, affordable approach. Because the Business Data Lake takes an augmentation approach, leveraging your existing EDW and related investments, we can make maximum use of any relevant technology or data that you already have, increasing your return on existing investments.

A light touch. You no longer have to create extensive inflexible GRC rules for every system and task. Instead you focus on risks, and take the appropriate level of action.

A general-purpose analytic resource. The BDL can be used for many types of analytics, extending the analytic capabilities of your business while complementing existing investments in security and EDW technologies.

All this means that implementing anomalous behavior analytics isn't simply another silo of information it becomes a core part of your companies information fabric, not only reducing risk but also enabling new insights and reports on the information it leverages.

Why work with Capgemini?

Capgemini have been a pioneer of this new approach of integrated information solutions. We have worked closely with the Pivotal data science team on the mathematical approaches required and our own internal data science team is collaborating with Pivotal on continually improving the approach. As a technology partner, we have deep expertise in Pivotal's technologies and data science skills backed up by our own dedicated Pivotal infrastructures in Europe and North America. We can provide anything from a rapid proof of concept (POC) to comprehensive integration, rollout and scalability services.

If you're interested in carrying out a POC, we can provide a working application in a few weeks. We take an anonymized sample of your data and show what insights you could gain by applying these techniques.

Once we have worked with you and identified the potential benefit for your organization, we can then scale the POC to provide a solution that meets wider line of business or full enterprise analysis, depending on your needs. We can provide a strategic roadmap to move from small-scale POC to a complete enterprise managed service, with analytics delivered in both real time and from long-term analysis.

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

Pivotal™

Find out more at
www.capgemini.com/bdl and
http://www.pivotal.io/big-data/businessdatalake

Or contact us at
bim@capgemini.com

About Capgemini

With almost 140,000 people in over 40 countries, Capgemini is one of the world's foremost providers of consulting, technology and outsourcing services. The Group reported 2013 global revenues of EUR 10.1 billion. Together with its clients, Capgemini creates and delivers business and technology solutions that fit their needs and drive the results they want. A deeply multicultural organization, Capgemini has developed its own way of working, the Collaborative Business Experience™, and draws on Rightshore®, its worldwide delivery model.

Learn more about us at
www.capgemini.com

MCOS_GI_PT_20140901

The information contained in this document is proprietary. ©2014 Capgemini.
All rights reserved. Rightshore® is a trademark belonging to Capgemini.