WHITE PAPER

Big Data @Work:

In the Fight Against Fraud, Businesses Are Finding the Best Defense Is a Good Offense

What's needed is a unified solution that works with relational as well as unstructured data to quickly and cost-effectively find hidden patterns and make predictions about troublesome activity

Introduction

Any business that has a secure network is vulnerable to unauthorized intrusion—and a potentially big hit to the bottom line. Security breaches cost U.S. companies an average of $11.6 million, an increase of 26 percent over the average cost reported in 2012, according to the Ponemon Institute. In the retail industry alone, more than 180 million records were compromised in 2011, with unauthorized transactions contributing nearly half of the more than $100 billion in losses stemming from fraudulent activities.

Threats to a company's intellectual property or financial systems can be internal as well as external. An average of $1.2 million is lost for every workplace fraud committed, and detecting a workplace fraud in North America takes an average of 4.2 years, according to KPMG. Physical assets are at risk as well. In the U.S. energy sector, illegal hookups and meter tampering contribute to more than $6 billion in losses annually. Even a seemingly small act such as a stolen credit card can have a significant impact on the card provider's brand reputation.

If those threats aren't daunting enough, cybercriminals continue to develop ever more sophisticated methods that are harder for IT organizations and their security teams to detect, and detect quickly. Anti-fraud teams, however, have an increasingly sophisticated weapon of their own to verify that people requesting access to secure information are who they say they are: big data.

Google's director of research, Peter Norvig, once said, "We don't have better algorithms. We just have more data." The concept that more data is better applies well to fraud detection, as IT teams explore ways to use data to engage in increasingly deeper levels of fraud analysis, all the way down to the transaction level.

All of this data, however, is putting a significant strain on existing data warehouses. In the ongoing fight against fraud, organizations are discovering that they require additional tools for storing, accessing, mining, and analyzing all this new data. The goal is a unified solution that can work with the existing relational data used to run the business today, and incorporate large quantities of new, less structured data to quickly find hidden patterns, discover new insights, and even make predictions about troublesome activity before it happens. Combining these types of data, using relational and nonrelational technologies, is one of the keys to what we call Big Data at Work.

ORACLE

intel

CIO
Custom Solutions Group

As the threats grow more sophisticated, companies are looking for solutions that analyze many more variables to uncover increasingly subtle anomalies, and do it quickly.

Becoming less reactive and more predictive

Fraud comes in multiple flavors, from multiple sources. Organizations have made significant progress on one front: protecting their networks from cyber attacks. But hacking is just one of many criminal behaviors that can be harmful to businesses and their customers. That's why IT is being asked to provide multiple layers of protections to reduce the risk of harmful activities such as identify theft or unauthorized use of credit cards.

At many organizations, an increased focus on fraud prevention means that ID management, business rules to flag suspect behavior (e.g., multiple, simultaneous purchases from different time zones), and blacklists (compilations of people who have previously committed fraud) have become checklist levels of protection. Traditional data warehouses can process most of this data.

But as the threats grow more sophisticated, companies are looking for solutions that analyze many more variables—from time windows to previous behavior to pre-authorization—to uncover increasingly subtle anomalies, and do it quickly. This is the potential of big data.

Organizations and their security teams must move from analysis of basic queries or business rules into a more advanced state that can sniff out trouble as it's happening. Predictive analytics leverage data to discover patterns and relationships among seemingly unrelated information to help companies make educated assessments of anticipated behavior—and take action to prevent that behavior. These types of real-time anomaly detection systems greatly increase the scale and types of data being collected, stored, and analyzed.

Several components must be in place to build and deploy a more sophisticated approach to fraud detection and prevention. The first key element is the data itself. To prime the big data pump, an organization will need access to suitable historical data, including customer transactions, call center logs, chat transcripts, emails and other structured and unstructured data, including social media. The second requirement is an infrastructure that can handle this greater variety of data. Combining an existing relational database with a nonrelational platform such as Hadoop is the best way to accommodate the rising tide of both structured and unstructured data *(see sidebar, page 3)*.

A third major requirement is advanced data-mining and statistical analysis tools that IT teams can use to identify patterns through different combinations of the data they are collecting. A product such as Oracle Advanced Analytics, for example, can be used to discover relationships hidden in data, including unstructured data in Hadoop environments. Oracle Data Mining, a component of Oracle Advanced Analytics, includes algorithms specifically designed to identify rare or anomalous records or detect flags in records that are believed to be "normal." Using these techniques, Oracle Data Mining can help businesses uncover suspicious submissions in expense reports, healthcare claims or tax forms, potentially saving large amounts of money in fraudulent claims.

Relational and nonrelational data: a powerful combination

Every organization is looking to make big data more useful. Until recently, most companies have extracted value from data by carefully selecting and standardizing the data collected based on predetermined relationships. This "run the business" approach is primarily about keeping existing systems and processes functioning properly. It is the world of the relational database.

But the increase in the variety, volume and velocity of data opens up the possibility to learn from the data in new ways. It's no longer as clear which data is most useful or what you can learn from it. By examining the data in a nonrelational environment, companies can form and test more hypotheses more quickly, resulting in new insights they would have missed otherwise. This "change the business" approach can lead to adjustments to existing processes and systems to achieve better results. Hadoop complements a relational database with its ability to store and process large volumes of nonrelational data.

The difference in these approaches boils down to one critical element: To run the business, you organize data to make it do something specific; to change the business, you take data as-is to figure out what it can do for you. Relational technologies excel at the first; nonrelational technologies are superior at the second. Together, these two approaches are more powerful than either is alone.

Turkcell gets a jump on calling card abuse

Turkcell İletişim Hizmetleri A.Ş. is a mobile communications provider in Turkey with more than 34 million subscribers. One of its product lines is a prepaid calling card that customers can use for a variety of financial transactions—a product type that is ripe for abuse. Anonymous network-branded prepaid cards are a tempting vehicle for money launderers, particularly since these cards can be used as cash vehicles—for example, to withdraw cash at ATMs.

Turkcell wanted to improve its ability to combat communications fraud and money laundering by introducing advanced analytical solutions to monitor key parameters of prepaid card usage and issue alerts or block fraudulent activity. This type of fraud prevention would require extremely fast analysis of the company's 1 petabyte of uncompressed customer data to identify patterns and relationships, build predictive models, and apply those models to even larger data volumes to make accurate fraud predictions.

The solution involved an Oracle Exadata Database Machine X2-2 HC Full Rack, powered by Intel® Xeon® processors. Data analysts can build predictive antifraud models inside the Oracle database, and then deploy these models into Oracle Exadata for scoring, using Oracle Data Mining. With this solution, predictive antifraud models run four hours faster than Turkcell's previous approach using a separate

Driving Big Data Performance

One of the biggest challenges with big data is finding an efficient, cost-effective and timely way to make it accessible to users. The faster your users are able to access and analyze big data, the quicker they can uncover new insights about the business.

Oracle and Intel are collaborating on engineered systems that are designed help customers acquire, organize, and analyze the flood of structured and unstructured big data coming into today's business. The engineered systems, which are powered by Intel® Xeon® processor E5 and E7 families include the following:

Oracle Exadata Database Machine: As a complete package of servers, storage, networking, and software, Oracle Exadata makes short work of intensive tasks—from online transaction processing (OLTP) to data warehousing. Oracle Exadata is scalable, secure, and redundant, and because it handles mixed workloads, it's an ideal platform for consolidating databases.

Oracle Big Data Appliance: The Big Data Appliance takes care of the heavy lifting associated with making big data available to end users. The extreme performance of this Intel-powered appliance enables rapid provisioning of a single system that's scalable, highly available, and optimized to transform massive amounts of data into a valuable asset for the business.

Oracle Exalytics In-Memory Machine: A best-in-class enterprise business intelligence (BI) platform, in-memory analytics software, and industry-changing Intel hardware all come together in this engineered system to provide extremely fast solutions for BI, modeling, forecasting, and planning. Faster analytics means more data analyzed, more users supported and more timely reports.

Oracle Exalogic Elastic Cloud: Whether your users are responding to customers, consolidating reports, or attempting to get the right product to market ahead of the competition, they need extreme performance and less complexity. Oracle Exalogic Elastic Cloud, a fully optimized, integrated platform for deploying and running business applications, enables users to access and get to work with their applications faster—improving business performance as well as the bottom line.

Relational and nonrelational data: A powerful combination

Every organization is looking to make big data more useful. Until recently, most companies have extracted value from data by carefully selecting and standardizing the data collected based on predetermined relationships. This "run the business" approach is primarily about keeping existing systems and processes functioning properly. It is the world of the relational database.

But the increase in the variety, volume and velocity of data opens up the possibility to learn from the data in new ways. It's no longer as clear which data is most useful or what you can learn from it. By examining the data in a nonrelational environment, companies can form and test more hypotheses more quickly, resulting in new insights they would have missed otherwise. This "change the business" approach can lead to adjustments to existing processes and systems to achieve better results. Hadoop complements a relational database with its ability to store and process large volumes of nonrelational data.

The difference in these approaches boils down to one critical element: To run the business, you organize data to make it do something specific; to change the business, you take data as-is to figure out what it can do for you. Relational technologies excel at the first; nonrelational technologies are superior at the second. Together, these two approaches are more powerful than either is alone.

analytics cluster. Analyses are performed more quickly because data does not have to be moved to a separate analytics cluster. Models can also be searched and queried using SQL inside the database, and Oracle Exadata can access raw data without summarized tables, further reducing analysis times. In the war against fraud, speed matters.

"We can analyze large volumes of customer data and call-data records easier and faster than with any other tool and rapidly detect and combat fraudulent phone use," said Hasan Tonguç Yılmaz, a Turkcell manager. Oracle Exadata also has the scalability to support Turkcell's rapidly growing data volumes.

Getting started

Each business must create a predictive model that is tuned to the behaviors of its own customers.

Successfully harnessing big data for fraud prevention encompasses two important steps:

1. Demonstrate clear ROI through a short-term success. A small-scale pilot that demonstrates business value will increase buy-in from key stakeholders, including senior management, and ensure proper funding for a larger roll-out.
2. Develop a long-term plan that serves as a baseline for expansion. A big-data strategy must give scale top consideration. All projects must be realistic about what's required, including support resources and maintenance of new software.

Summary

Data is arguably an organization's most valuable asset, because it represents the entire history of the business and its interactions with customers. As organizations do more business through digital channels, they need to go beyond simply defending their data against cybercriminals; they need more advanced methods to rapidly predict potential fraudulent activity and take steps to prevent it.

There's a lot at stake—not just the costs of fraudulent activities, but the very reputation of the brand in the eyes of its customers. As criminals and other bad players look to exploit vulnerabilities in a business's defenses, it's time for IT organizations to be more proactive. An integrated big data solution that leverages relational and nonrelational technologies will enable security teams to more effectively predict—and prevent—fraudulent activities before they do real damage to the enterprise.

For more information, visit www.oracle.com/bigdata and www.oracle.com/goto/nosql

1 2013 Cost of Cyber Crime Study, Ponemon Institute, October 2013, http://www.networkworld.com/community/node/83988

2 2011 True Cost of Fraud Study, LexisNexis, http://lexisnexis.com/risk/downloads/whitepaper/tcf_2011.pdf

3 Who Is the Typical Fraudster?, KPMG, http://www.kpmg.com/US/en/IssuesAndInsights/ArticlesPublications/Documents/who-is-the-typical-fraudster.PDF

4 Energy Theft Goes Global, The Energy Collective, February 2013, http://theenergycollective.com/sbattaglia/185556/energy-theft-goes-global

5 Why Google is Offering 411 Service, O'Reilly Radar, April 2007, http://radar.oreilly.com/2007/04/why-google-is-offering-411-ser.html