Objectivity

Solution Brief

How Discovering Data Relationships Can Fight Cybercrime

The Business Problem

Every day that cyber threats go undetected results in the potential for more data theft, creating increased long-term repercussions to businesses. In order to minimize the damage from cybercrime, organizations need the ability to quickly identify abnormal activity on their networks so that they can quickly isolate the problem and react accordingly. They need the ability to access historical data and analyze it to uncover patterns, so that they will be able to more quickly discern when unusual activity is occurring. This involves advanced relationship and pattern discovery processes.

While “known” threats can often be identified by common anti-virus software, firewalls, and event management tools, “unknown” threats take new forms, and may not be immediately spotted based on common queries.

The Technical Challenge

Organizations today make use of historical data and logs to recognize patterns and connections within their data, analyzing archival data alongside streaming data to quickly ascertain discrepancies and potential threats. This typically involves analyzing the following:

- **Signatures** of past breaches to identify known instances of cyber threats.
- Multiple data points (e.g. time of activity, frequency of activity, location) and how they relate to **historical norms** for both the individual user and past trends.
- Relationships between anomalies in **social networks** associated with key individuals in the organization.

Systems set up for these include a complex data ingest layer–where streaming data is transformed–and a graph-like storage layer where this data and the relationships between various transactions can be persisted, then rapidly and continuously queried.

This can often be a challenge, as the volumes of data that must be consumed and analyzed continue to increase. The creation of an ingest layer that can consume, transform and store streaming data while creating and maintaining information about the relationships between transactions becomes very complicated. At many times, it becomes a stumbling block.

Data Breaches on the Rise

- ***783:*** The number of reported breaches at U.S. organizations in 2014, a record high
- ***$2 trillion:*** The global cost of breaches expected by 2019 [1]

A Cautionary Case Study: Ashley Madison

- ***32 million:*** Users of the online dating service that were hacked in 2015
- ***15,000:*** U.S. government workers exposed, implicating national security
- ***$760 million:*** Damages claimed in a class action lawsuit [2]

The Need for Real-Time Response

The average time taken to discover data breaches:

- ***98 days*** for financial firms
- ***7 months*** for retailers [3]

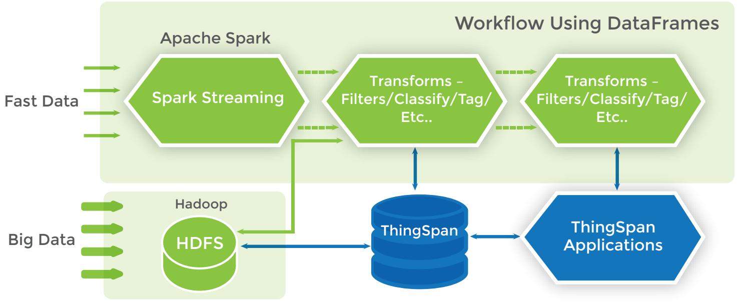
[1] 2015 report from The Economist
[2] 2015 report from CNN
[3] 2015 report from Ponemon Institute

In addition, most of the graph stores available today are not designed to scale to multi-billion nodes and edges while supporting billions of transactions that need to be analyzed and queried per day: this is the level of performance and scalability needed to identify emerging threats quickly enough to stop them before significant damage is done.

As a result, most organizations rely on solutions based on a custom-built ingest layer feeding into a graph database to maintain relationships, which neither scale nor support required query times. The resulting solutions suffer from limitations that include:

- Complex ingest path solutions, which are **expensive** and difficult to maintain.
- Inefficient and expensive hacks like pre-computing sub-graphs for acceptable query times, and therefore **limited functionality**.

ARCHITECTURE DIAGRAM


Workflow Using DataFrames
Apache Spark
Fast Data
Spark Streaming
Transforms - Filters/Classify/Tag/ Etc..
Transforms - Filters/Classify/Tag/ Etc..
Hadoop
Big Data
HDFS
ThingSpan
ThingSpan Applications

The ThingSpan™ Solution

ThingSpan™, Objectivity's Fast Data solution platform, is integrated with **Hadoop** and **Spark** to give organizations the capability to build a fully supportable cybersecurity solution. It does this by enabling organizations to ingest, transform and consume massive and varied data streams to create and persist complex, scalable graph structures. These structures can operate at petabyte scale and efficiently support complex, continuous queries.

ThingSpan leverages open-source tools by supporting the Hadoop and Spark ecosystem atop a high-performance, distributed graph database purpose-built for relationship and pattern discovery. It runs natively on top of HDFS as a YARN application while using Spark for workflow and data transformation. It is also designed to support streaming systems based on Kafka, Flume and other distributed messaging tools for streaming data. Integration with Spark via DataFrames allows ThingSpan to ingest streaming data while maintaining and persisting relationships as first-class logical models.

This model allows for enriched and transformed data to simplify the support of complex, multi-dimensional queries associated with cybersecurity applications and analytics. With its relationship-oriented approach to information fusion involving fast, streaming data and static, historical and transactional data, ThingSpan delivers optimal intelligence to fight cybercrime. Now organizations can achieve business insights from Big Data and real-time streaming data with a high degree of efficiency at scale, thereby preventing future security breaches.

Objectivity

Objectivity, Inc. is a pioneer in high-performance distributed database platforms that power mission-critical applications for the most demanding and complex data sources in the enterprise. Objectivity enables organizations to accelerate time-to-value of their data assets at scale by enriching Big Data with Fast Data. With a rich history of serving Global 1000 customers and partners, Objectivity holds deep domain expertise in fusing vital information from massive data volumes to capture new revenue opportunities, drive competitive advantages, and deliver better business value. Objectivity is privately held with headquarters in San Jose, California. Visit http://www.objectivity.com to learn more.

THINGSPAN™

ThingSpan is a purpose-built Information Fusion platform designed to accelerate the time-to-production of Industrial Internet of Things (IoT) applications. It is a massively scalable distributed platform designed specifically for the complex issue of extracting actionable insights from Fast Data and Big Data sources. It is architected to integrate with major open source Big Data technologies, such as HDFS and Apache Spark. ThingSpan leverages Objectivity's core object data-modeling technology and the company's rich experience in building fusion solutions.