Big data: the next frontier for investigations and regulatory response

By Todd Marlin

Those who have worked with electronically stored information for more than a decade or so no doubt remember the hand-wringing days of trying to extract relevant data for litigation purposes from what seemed to be an ocean of gigabytes. Just as we became comfortable with the concept of gigabytes worth of storage, our initial reaction to the exponential growth of data then was “what in the world is a petabyte?”

We long for the days when we faced a paltry 20 million emails, as in the 1999 case of United States v. Philip Morris — a seminal e-discovery case that was among the first to tackle large volumes of data. Cases like this were phenomenal for their time, but are now a matter of routine.

Today, an hour’s worth of business for a typical big-box retail chain can create millions of transactional records. The entirety of data from the private sector doubles every 14 months. So-called “big data” is no longer a phenomenon. It is a reality we wrestle with in the everyday course of business to a degree we never conceived, and it touches myriad worlds — not the least of which are those of anti-corruption and fraud.

To understand the challenges and opportunities presented by this seemingly ungovernable deluge of data, we must first understand what big data is. Since the advent of consumer-level networking and the internet (machines talking to machines), massive volumes of data across disparate systems have surged.

These data have grown so large and unwieldy that it has become a daunting prospect for companies to capture, curate and properly leverage them to drive revenue, prevent loss or respond to an episodic event when facts are in question — so daunting, many companies prefer to keep their collective heads buried in the sand.
What exactly is so scary? Consider that when your organization leaves the league of petabytes in storage and moves to exabytes (that's a 10^18 bits, or about one thousand petabytes), you are then working at an organization that stores more data than the entirety of human civilization until about 20 years ago.

It can be hard to wrap one's head around concepts like that and even harder to devise a plan to handle, target and interrogate that data when the need arises. Our team has made a science of helping clients to leverage these data, making them see the worlds of information not as burdens but as chances to curb waste.

Never before have companies had the chance to so thoroughly understand their customers. Madison Avenue advertising firms made fortunes in the past century purporting to understand how to engage target demographics.

Today, that information is in the hands of corporations. Likewise, the opportunity to spot fraud before it gets off the ground has never been greater. But to accomplish either, the right tools are needed.

Once it was enough to operate on platforms like Oracle or SQL Server, but in the age of big data, clickstream, sensor data, log files, geospatial mobile data and social network interactions are all in play — and comprise a dataflow too heavy for old-school solutions. Today, products like Greenplum (EMC), Apache Hadoop and Teradata are required to support and scale heavy virtual warehouses. Open-source methodologies, such as MIKE2.0, also exist.

However, newer technology alone is not the answer. Subject matter experience is necessary when dealing with a niche specialty such as fraud, anti-bribery and anti-corruption, investigations, and litigation to truly leverage big data. Companies use teams like ours to change and design these systems to make them effective on an ongoing basis or in reaction to litigation, investigations or regulatory events.

To do this we go beyond traditional rules-based tests and use text analytic tools to allow the data to speak for itself. Many big data analytic platforms are also being developed; some of these include HP Vertica, Palantir and EMC Greenplum.

In short, big data is everywhere. From transactional systems to network security and social media, it is the all-encompassing backbone of corporate America. Accounting and related supporting systems all deal with big data.

In this new world, if you are responding to litigation issues or regulatory inquiries, are engaged in an investigation or are trying to prevent one, you will need to deal with big data. The key to tackling this new frontier effectively is to leverage new technology in conjunction with subject matter experts to laser focus the technology on the data at issue.