DATAART

ITBUSINESSEDGE

Isolating Sensitive Bits of 'Big Data' Key to Compliance

“Big Data” is the latest buzzword to sweep through enterprise IT, and — as you might expect — it is projected to affect everything your company does. Not surprisingly, the “Big Data” conversation is moving toward compliance, or should we say “Big Compliance.”

A quick survey piece at **Wall Street & Technology**places **Governance and Risk Reporting on a list of 10 Big Data ideas** that will impact the financial sector in light of the Dodd-Frank Act and other stiffening regulations. Well, yeah.

The governance and compliance challenges presented by Big Data and its core technology framework, Apache Hadoop, are as sizable as they are self-evident. Hadoop is a Google-derived technology that enables applications to run in bite-sized chunks on clustered commodity hardware, and when appropriate to take chunks of data, both structured and unstructured, with them for the ride. The net effect is that apps no longer need to rely on RDBMS software or standardized data structures to mash away at all that data coming from all over the globe.

Trick is, strictly enforced data structures are pretty darn handy when it comes to access, visibility and other data management issues. Remember data warehouses? — there was a reason folks were willing to go to all that pain in getting their info into one place, with one set of rules.

In the Wall Street & Technology piece, an IDC analyst is quoted as saying:

> *… big data solutions that support evolving business and regulatory requirements by maintaining an ecosystem of large data sets will become invaluable in months or years from now.*

Again, well yeah.

Our Mike Vizard this morning reports that **at least one compliance vendor has introduced a version of its toolsets adapted to Big Data rules enforcement**. The CEO of the product vendor notes that an underlying risk of accessing unstructured (or “raw”) information via Big Data technologies is that the sensitivity of that data may not be immediately apparent.

A **Dark Reading post** from earlier this year discusses this issue at length, with one security vendor executive warning that companies are overlooking the importance of finding parcels of data subject to regulations like PCI. Jon Heimerl, director of strategic security for Solutionary, is quoted as saying:

> *Mainly, big data stores are leading organizations to not worry enough about very specific pieces of information.*

The Dark Reading piece does offer the solace of noting that lawmakers tend to drag their feet when it comes to regulating new technologies, and so IPSec pros have time to get their heads around Big Data’s compliance implications. It also closes with advice to begin segregating, or “siloing,” what your company perceives as vital information in the Big Data soup for future compliance rigors.

For those firms who have already tried to tackle a Big Data compliance layer, one common pitfall has been **treating the information as though it lives in a relational DB**, according to DataArt exec Oleg Komissarov in a guest column at Wall Street & Technology. Some financial firms had written software that runs atop memory caches as though it was hitting a DB — not good.

So, it appears that exactly how compliance will be realized for Big Data is still an open question. The consensus is to just adhere to best data practices and know that discrete compliance rules are coming — probably right as the next “Big” trend hits enterprise data management.