BIGDATA SPECIAL

CIO Review

The Navigator for Enterprise Solutions

OCTOBER - 30 - 2015 CIOREVIEW.COM

Company of the Month:

Ty Moser, Founder, President & CEO, Moser Consulting

In My Opinion:
ANDRE FUETSCH, SVP, AT&T

CIO Insights:
RANDY SLOAN, SVP & CIO, SOUTHWEST AIRLINES

TransUnion: Smarter Decisions through Big Data

Jim Peck, CEO & President

$ 15US SE 12
0 74470 91626 0

CIO REVIEW
44790, S Grimmer Blvd.
#202, Fremont, CA-94538

Overcoming Security Challenges Associated with Big Data

By John Armstrong, CMO, Zettaset Inc

The amount of digital data in the universe is growing at an exponential rate, doubling in size every two years. By 2020 the digital universe, the data we create and replicate annually, is estimated to reach 44 zeta bytes, or 44 trillion gigabytes. This massive amount of data is referred to as "big data". Data is now created by multitudes of devices—everything from smart power meters to cell phones to environmental sensors. And there appears to be no limit to its growth.

Big data has caught the attention of big businesses as a way to out-market and out-sell peers. By capturing and carefully analyzing buyer data, companies can determine purchasing patterns and consumer sentiment, and create offers that are precisely targeted to individual customer profiles. In a recent report, McKinsey estimates that a retailer embracing big data has the potential to increase its operating margin by more than 60 percent, and cites many examples where companies have taken market share from competitors in verticals such as financial services and insurance.

The technology most closely associated with big data is Hadoop. Using commodity hardware and a distributed computing architecture consisting of multiple servers networked together, Hadoop can cost-effectively achieve tremendous scale. In terms of IT efficiency, Hadoop can bypass the need to use expensive data warehousing solutions from established vendors like Oracle and Teradata and thus has the potential to deliver impressive OpEx efficiencies.

Hadoop arose from efforts at Yahoo and Google driven by their need to inexpensively store and process petabyte levels of data. Both companies were struggling to manage the massive amounts of data they were collecting from the Internet. The vast majority of the data was public and there was no compelling reason to address security. Public URLs and IP addresses require no data protection. As a result, security was of little consideration when the Apache Hadoop open source project was launched in 2006.

The advent of Hadoop ushered in a new industry currently dominated by Cloudera, Hortonworks, and MapR. So far, Hadoop market growth has been steady but relatively modest. According to a report released in September 2015 by Relato, the top three Hadoop vendors together count less than 1,700 customers. While not large, this number is growing

and consists primarily of enterprise organizations. This makes sense, since it is the largest corporate and government organizations that work with huge data stores. And secondly, only well-capitalized organizations have the resources necessary to get complex Hadoop projects off the ground.

With major security breaches and fraud incidents making international headlines, organizations are taking steps to address the growing problems of advanced persistent threats, fraud, and insider attacks. Big data stores - whether Hadoop, NoSQL, or traditional relational databases - are candy to smart cyber criminals looking to make a big score by stealing sensitive information assets, intellectual property, credit card numbers, and customer databases.

When it comes to securing big data, however, traditional security technologies may not always be up to the task. Hadoop is built on a distributed computing architecture, with "clusters" consisting of multiple servers or "nodes". These can easily scale up into the hundreds. Relational databases typically have a more monolithic, centralized architecture. Therefore, securing Hadoop requires the ability for a security technology to scale as more nodes are added to a cluster, and as the number of clusters grows. To further complicate the challenge, data is moving between nodes as the cluster reconfigures itself to accommodate more data and job processing. This can make protecting data feel like shooting at a moving target.

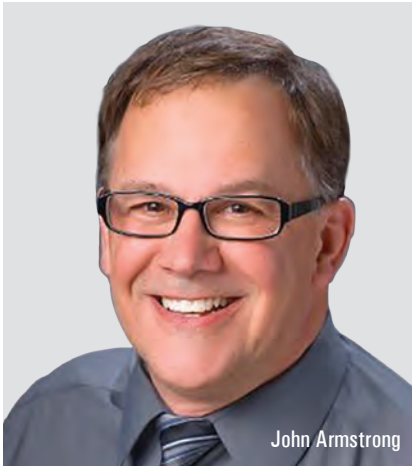
Technology vendors with specialized domain expertise in access control, authentication, and cryptology are rising to the challenge with solutions that take a more focused approach to securing data in Hadoop and NoSQL big data environments. Organizations wishing to secure big data stores should consider the following criteria when evaluating and selecting a security solution.

• Can it easily fit into existing IT security frameworks? Some security offerings designed for Hadoop, such as open source encryption, lack standards, interoperability with key management (KMIP) and hardware security modules (PKCS#11). This makes it difficult to fit into existing enterprise security frameworks where enterprises have already made investments in key managers and HSMs. Look for a solution that meets OASIS standards criteria.

“With major security breaches and fraud incidents making international headlines, organizations are taking steps to address the growing problems of advanced persistent threats

• How compatible is it with various Hadoop distributions NoSQL databases? For example, some security solutions, such as those from Cloudera, may not work well with other Hadoop distributions. This can limit your flexibility if you are not committed to a single Hadoop vendor.

• Can it integrate with multiple file systems in the data center? Integration


John Armstrong

with HDFS is table stakes in the Hadoop environment. A solution that can also support GPFS and other file systems is a bonus, and can save headaches down the road.

• Is it cluster-aware and optimized for a distributed computing architecture? Many access control and encryption solutions from long-standing security vendors were never designed for newer distributed computing architectures like Hadoop. As a result, unexpected performance issues may arise when deploying legacy security solutions in data centers where cluster node counts are increasing to meet expanding data requirements.

• Can it provide support for legacy relational databases as well as newer big data stores? This is a tough one, because security solutions designed for Hadoop/NoSQL environments may not work well in monolithic data architecture. Significant in-house testing may be required to determine suitability in your particular hardware and application environment. There is certainly value in being able to deploy a single security technology across multiple database environments. Remember: one size may not fit all. CR