SOLUTION BRIEF

BIG DATA SECURITY

Get maximum value and insight from your Big Data initiatives while maintaining robust data security

TRACE3

THE CHALLENGE

More and more companies are finding that Big Data strategies can pay great dividends and yield real competitive advantage. But in their eagerness to adopt Apache™ Hadoop® and other Big Data platforms, many businesses are ignoring the potential security risks.

Unlike the traditional database world, where a hacker would need to crack multiple databases and find a way to join the tables, Big Data analysis conveniently puts all the eggs in one basket. The "crown jewels" are now stored in clear text that any systems engineer with administrator permissions can easily pull into a spreadsheet. This has been a problem at the highest levels—the well-publicized NSA leak involving whistleblower Edward Snowden was a byproduct of Big Data analysis, for example.

Traditional access control isn't enough, and using standard data encryption methods can actually limit the value you gain from your Big Data initiative. This is because encryption obscures the formatting of and relationships between data points, making it impossible to put the "big picture" together without unencrypting the data.

Forward-thinking organizations are taking an approach called tokenization to protect sensitive data points while still gaining maximum insights from Big Data.

HOW TOKENIZATION WORKS

Tokenization is different from encryption in that it is based on randomness, not on a mathematical formula. It uses codebooks instead of cryptographic keys, replacing sensitive data (such as credit card numbers) with random "code text" to mitigate the chance that the data can be used effectively for malicious purposes. The token cannot be discerned or exploited, since the only way to get back to the original value is to reference the lookup table that connects the token with the original value. There is no mathematical formula, just a referencing system that only the data owner can decode. No one else can make the connection between data points and what that data describes.

A token looks like the original value in data type and length, enabling it to travel inside most applications, databases, and other components without modifications, resulting in greatly increased transparency. This also reduces remediation costs to applications, databases, and other components where sensitive data lives, because the tokenized data will match the data type, length, and format of the original. Data is better protected because of the limited need for de-tokenization/ data in the clear (due to the higher level of transparency).

Tokenization can be used to safeguard any type of personally identifiable information (PII), and is often used for PCI compliance.

PROTEGRITY "VAULTLESS" TOKENIZATION

Protegrity has developed a patent-pending, Vaultless Tokenization process that secures the data itself without storing sensitive data or tokens in a centralized "vault." Vaultless Tokenization eliminates performance and scalability bottlenecks, allowing customers to scale with the platform—the more nodes, the better the performance. This provides the ability to process as many as 200,000 tokens per second without compromising security.

Protegrity token services can be deployed in a distributed environment or in a central topology, without requiring replication between the token servers. Having the choice to deliver a distributed or centralized Vaultless Tokenization solution can optimize performance and security.

BIG DATA PROTECTION FOR HADOOP

Going well beyond traditional access controls and volume encryption, Protegrity offers a comprehensive data security path with both coarse and fine-grain data security options for Apache Hadoop, Teradata® Portfolio for Hadoop, Cloudera®, Pivotal®, Hortonworks®, MapReduce, and

other Big Data platforms. With the Protegrity Data Security Platform, only security officers have the ability to grant someone the authority to view sensitive data in the clear. System administrators and engineers can thus be prevented from seeing sensitive data or granting sensitive data access to others. This will not, however, prevent these specialists from performing their jobs.

Protegrity Big Data Protector gives Hadoop users top-to-bottom data protection, from the file to the application. It's a complete solution for not only securing the data itself, but also for providing role-based protections and audit trails that go beyond the simple access controls and network segmentation available in Hadoop today.

By utilizing Vaultless Tokenization on cluster nodes for individual data elements, Protegrity Big Data Protector offers infinite scalability and leverages the parallel processing power of Hadoop. It provides central data security policy management for access control and secures all sensitive data in Hadoop in any state—at rest in the Hadoop Distributed File System (HDFS); in use during MapReduce, Hive, or Pig processing; and in transit to other enterprise data systems. The actual data is protected from external and internal threats, and users and business processes can continue to mine the secure data for transformative decision-making insights.

Protegrity Big Data Protector protects any sensitive file stored in the HDFS with volume-level strong encryption, and any sensitive data item stored within a file using tokenization. For unstructured data, Protegrity Big Data Protector provides transparent data protection to HDFS layers—including the first integrated HDFS encryption that utilizes native HDFS functionality (Intel Project Rhino)—thus protecting any files stored within the Hadoop cluster. For structured data, Protegrity Big Data Protector also extends MapReduce, Pig, and Hive with data protection as well as Protegrity's tokenization technology that replaces sensitive data with "fake" data while also providing data type preserving encryption.

BENEFITS OF PROTEGRITY BIG DATA PROTECTOR

- Security during Big Data analysis with minimal performance impact
- Complete protection for Hadoop, Cloudera, Teradata Portfolio for Hadoop, Hortonworks, MapReduce, BigInsights, and other Big Data platforms
- Protection in HDFS, MapReduce, Hive, and Pig
- Extensible throughout the enterprise Big Data ecosystem (not just a point solution)
- Top-to-bottom, granular protection, from files to applications
- Ability to restrict users from accessing sensitive data
- High performance and infinitely scalable, with low total cost of ownership

DON'T WAIT—SECURE YOUR DATA NOW

Taking a proactive stance on Big Data security will greatly help your organization down the road. Trace3 recommends using a tokenization solution such as Protegrity to meet compliance requirements while confidently getting the most from your data. Your data will grow, and the regulatory landscape will only get more intense. It's easier and cheaper to secure your data now than to clean up a mess later.

Beyond regulatory requirements, however, avoiding security breaches requires an honest risk assessment. Compliance does not equal security—companies should assess the true cost of a data breach and take proactive measures with their Big Data initiatives. Trace3 can work with your company to determine what data needs to be tokenized, and help you deploy and get the most value from Protegrity Big Data Protector for Hadoop.

ABOUT THE AUTHORS:

Scott Mellegaard

Scott has been involved in all aspects of IT for the last 20 years and is certified in several various technologies. As a founder of several technology startups, and now the Business Data Intelligence (BDI) division within Trace3, Scott brings a unique perspective on how all the pieces of technology orchestrate together and provide true value to the business.

Nick Durkin

Nick Durkin possesses over ten years of data center architecture knowledge combined with a wealth of certified training, Nick also owns over three years of Hadoop and Big Data experience, architecting systems and building big data applications for not only the Federal government but also the world's largest financial institutions. Nick has held the position of lead architect on the Department of Homeland Security's FIVICS initiative as well as developed many of the anti-fraud applications currently in use amongst the aforementioned financial institutions.

ABOUT TRACE3'S BIG DATA INTELLIGENCE TECHNICAL BRIEFS

The Trace3 Big Data Intelligence (BDI) team finds and vets innovative solutions that apply to common, real-world use cases. Once a solution has been identified and gone through our technical vetting, it is installed in our BDI Innovation Labs for extensive integration testing with other complementary technologies. A Technical Brief document is developed with the participation of the solution providers involved.

FOR MORE DETAILED TECHNICAL INFORMATION OR TO RECEIVE A DEMO OF THIS SOLUTION, CONTACT THE TRACE3 BIG DATA INTELLIGENCE TEAM.

BDI@TRACE3.COM | WWW.TRACE3.COM/BIGDATA

Create the first-mover advantage and transformational clarity that ensures a competitive edge in today's marketplace. For more information, visit **www.trace3.com**