Voltage security

SOLUTION BRIEF

Voltage Enterprise Security for Big Data

Momentum behind Big Data is growing with recognition of the competitive advantage for companies who successfully harness Big Data versus those who delay. "Data-driven decision-making enables a statistically and economically significant lead in profitability and productivity"*, according to the Harvard Business Review. Enterprises need to enable access to data for analytics – in-house, off-shore, in the cloud, and with partners, vendors, and others – to extract maximum value from information. Use of the Apache Hadoop architecture is growing dramatically. However, lack of a comprehensive data protection strategy can be a major obstacle to taking Hadoop projects live. Whether only beginning or well underway with Big Data initiatives, companies need data protection to mitigate risk of breach, assure global regulatory compliance and deliver the performance and scale to adapt to the fast-changing ecosystem of Hadoop tools and technology. Voltage SecureData™ for Hadoop delivers the data protection strategy organizations need to deploy Big Data initiatives for competitive advantage.

A Data Protection Strategy to Enable Big Data Initiatives

Business insights from Big Data analytics promise major benefits to enterprises – but launch of these initiatives also presents massive potential risks. Architectures like Hadoop can aggregate structured, semi-structured and unstructured data, perform parallel computations on large datasets, and continuously feed that store to enable data scientists to see patterns and trends. Because companies can now cost-effectively hold massive amounts of data for analysis, potentially representing years of information, Hadoop systems can increase the risk and magnitude of data breach. Sensitive data can be exposed, and thus violate compliance and data security regulations. Aggregating data across borders can break data residency laws. So, finding ways to effectively secure sensitive data, yet enable analytics for meaningful insights, is a top obstacle to implementing Big Data capabilities broadly throughout the organization.

Companies shopping for Hadoop solutions are seeing the need for a comprehensive data protection strategy. Across industry and government, organizations need data protection solutions with these capabilities:

- Security – protect data from any source, of any format, before it enters Hadoop; enable access to internal and external users, with protection of sensitive data that maintains usable, realistic values for accurate analytics and modeling on data in its encrypted form.
- Compliance – securely capture, analyze and store data from global sources, and ensure compliance with international data security, residency and privacy regulations. Address compliance comprehensively, not system-by-system.
- Performance and extensibility – integrate data security fast, with quick implementation and an efficient, low maintenance solution that won't degrade performance and will scale up. Leverage IT investments by integrating with the existing IT environment and extending current controls and processes into Hadoop.

Voltage SecureData for Hadoop – Security for Big Data Analytics

Voltage SecureData™ for Hadoop enables companies to respond immediately to the business need for Big Data analytics by implementing high performance data security with extensibility, scale and adaptability to Hadoop and other Big Data technologies. By uniting market-leading Voltage Format-Preserving Encryption™, data masking, Voltage Secure Stateless Tokenization™ technology and Stateless Key

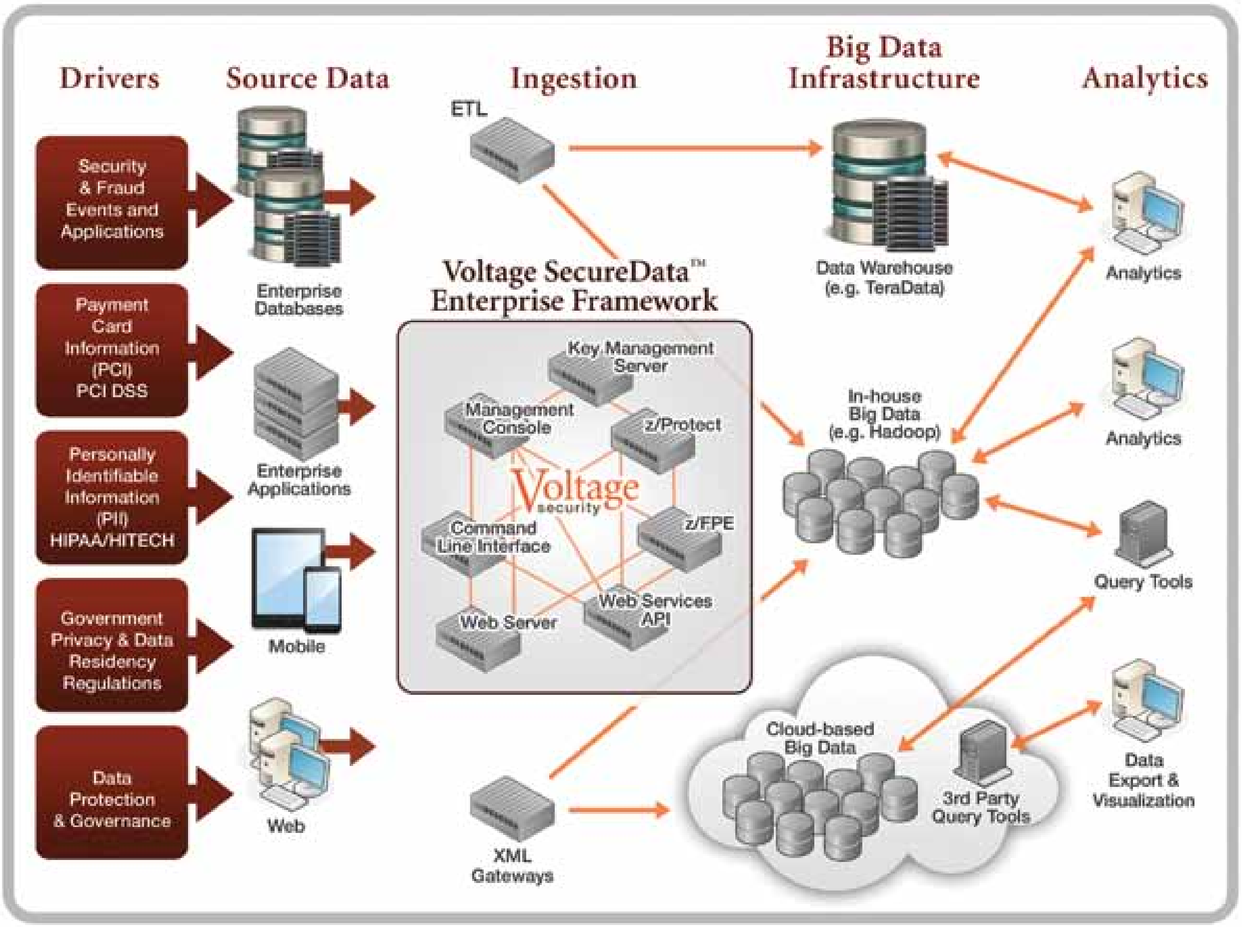
Management in a single solution, Voltage SecureData for Hadoop delivers comprehensive compliance with national, industry and international privacy regulations. By ensuring data-level protection that reduces the risk of inappropriate access and attack, it enables the secure access and use of data across the extended enterprise.

The unique values and capabilities delivered by SecureData for Hadoop mean this: data protection for Hadoop can be rapidly implemented to reduce security and compliance risks, and help propel maximum return on information[1].

Solution Benefits

- Protect data on-the-fly into Hadoop; easily extend data security enterprise-wide
- Provide secure access for data analytics internally, externally, and in the cloud
- Ensure usable data value for analytics, but no value for attack!
- Analyze global data in compliance with international regulations
- Rapidly integrate high-performance data security into existing IT environments
- Dramatically reduce maintenance by eliminating the need to constantly back up key stores
- Grow Big Data security with true horizontal scaling – responsive to business demand growth
- Flexibly adapt data security to the fast-growing ecosystem of the newest Hadoop tools and technologies.
- Integrate with tools that run across the enterprise, in z/OS mainframes, Linux, UNIX, and Windows environments, so that data can be protected upon capture, and flow into Hadoop without the need to decrypt and re-encrypt.

Enterprise Security for Big Data


Drivers
Source Data
Ingestion
Big Data Infrastructure
Analytics
Security & Fraud Events and Applications
Payment Card Information (PCI) PCI DSS
Personally Identifiable Information (PII) HIPAA/HITECH
Government Privacy & Data Residency Regulations
Data Protection & Governance
Enterprise Databases
Enterprise Applications
Mobile
Web
ETL
Voltage SecureData™ Enterprise Framework
Key Management Server
Management Console
z/Protect
Voltage security
Command Line Interface
z/FPE
Web Server
Web Services API
XML Gateways
Data Warehouse (e.g. TeraData)
In-house Big Data (e.g. Hadoop)
Cloud-based Big Data
3rd Party Query Tools
Analytics
Analytics
Query Tools
Data Export & Visualization

[1] "Big Data: The Management Revolution", by Andrew McAfee and Erik Brynjolfsson, Harvard Business Review, October 2012

Capabilities of SecureData for Hadoop

Security with Format-Preserving Encryption™ (FPE): Structured, semi-structured or unstructured data can be encrypted at source and protected throughout the data life-cycle, wherever it resides and however it is used. Protection travels with the data, eliminating security gaps in transmission into and out of Hadoop and other environments. FPE enables data de-identification to provide access to sensitive data while maintaining privacy and confidentiality for certain data fields such as Social Security Numbers that need a degree of privacy while remaining in a format useful for analytics. Unlike competitive solutions, SecureData is not encrypting at the Hadoop file system (HDFS) layer, rather protecting only the fields that need protecting.

- Protects data aggregated on its way into Hadoop, or selectively inside MapReduce jobs
- Preserves format so protected data has an appearance that matches the original data (e.g. credit card, Social Security Number, national ID formats) – a vital characteristic for analysis to operate properly
- Provides reversibility, so original data can be securely retrieved per business policy controls
- Delivers standards-based, patented and NIST-recognized security, with published security proofs

Compliance: For projects that must meet compliance with industry and international data security, residency and privacy regulations, patented FPE and Secure Stateless Tokenization techniques enable dramatic reductions in audit scope, costs and complexity. Big Data initiatives often aggregate data from global sources crossing national boundaries. With Stateless Key Management, data can be analyzed in protected form in one jurisdiction, and data decryption/de-tokenization applied in another jurisdiction where specifically permitted.

- Removes international data residency legal restrictions to cross-border data aggregation
- Puts Hadoop out-of-scope for Payment Card Industry Data Security Standard (PCI DSS)
- Assures compliance for Big Data applications with industry, data security and privacy regulations applicable to Personally Identifiable Information (PII) and Personal Health Information (PHI)

Performance: Many applications function with FPE-encrypted or tokenized data, with no need for decryption. This is due to the preservation of format and referential integrity, and the ability to leave selected characters "in the clear" so the data can be used but not compromised. For example, purchase behavior could be analyzed by month and year, with obfuscation of the particular day and time records for each purchase (2011:10:AD 1Y:33:9B). Compute-intensive applications usually perform their analytics on protected data, and then decrypt the final result set only if needed. Data can be FPE-encrypted at the source then simply imported into Hadoop, and used without re-encryption. This is extremely efficient and decreases the ingestion time to get protected data into Hadoop because there is no need for an encryption step at that point.

- Encrypt terabytes of new live data on the fly into thousands of Hadoop nodes in parallel
- Perform multimillion encryption operations-per-second on large clusters
- Support massive data volumes as well as business demand growth with linear scalability

Integration: Voltage SecureData for Hadoop can immediately integrate with virtually any application, ranging from decades-old custom applications to the latest enterprise applications. SDKs/APIs and command line tools enable encryption and tokenization to occur natively on the widest variety of platforms, including Linux, mainframe and mid-range. APIs enable broad integration into portfolios including ETL, cloud, databases and applications – and Hadoop, with native on-node cluster-wide data-masking, encryption and decryption.

- Adds on-the-fly data protection to ingestion scripts and ETL tools (such as those from Talend, IBM and Informatica) using Voltage's Web Services, Java or C APIs. Voltage provides "user defined functions" (UDFs) that plug in to HIVE as well as Teradata. These UDFs can also be adapted to run in PIG, IBM Big Insights, and a variety of other analytics tools
- Integrates into query languages such as PIG and Hive, and also into MapReduce jobs
- Provides multiple integration points including network level XML gateways and ETL tools for batch operations

Extensibility: Voltage Stateless Key Management is vital for global organizations. With a highly available, distributed architecture, it provides keys automatically without the key storage or database management challenges associated with legacy approaches. Voltage Stateless Key Management can be linked to existing Identity Management infrastructure including roles and groups. Permission to decrypt or de-tokenize can be assigned on an application or user basis, and can be managed through external enterprise directories, taking advantage of existing identity management solutions to simplify user management. The result is role based access to data at a data field level, mapping directly to enterprise data access rules and policies.

- Enables extension of enterprise controls into Hadoop
- Eliminates the need for dedicated IT headcount for key management
- Removes the need to constantly backup key stores and dramatically reduces complexity of maintenance
- Provides flexible and multi-layered framework for managing authentication and access control for highly restricted and sensitive content

Conclusion

Voltage SecureData for Hadoop delivers comprehensive data security enterprise-wide. This solution enables technical stake-holders to be responsive to the business need to immediately start extracting more value from information in Big Data initiatives, without introducing risk to sensitive corporate information or regulatory compliance. For enterprises now launching a Big Data program, Voltage SecureData for Hadoop can be up-and-running in weeks to secure sensitive data before it enters Hadoop environments. Voltage SecureData for Hadoop removes the top obstacles to moving forward with Big Data initiatives, and enables integration of Big Data analytics and insights broadly, throughout the extended enterprise.