
HADOOP DATA PROTECTION

At-Rest Data Encryption and Secure Key Management for Enterprise Data Hub



Securing The Enterprise Data Hub

AT-REST ENCRYPTION

High-performance data-at-rest encryption for Hadoop that meets standards for compliance and safeguarding PII

MASSIVE SCALABILITY

Node-based encryption for the fastest, most scalable security on Hadoop

POWERFUL, FLEXIBLE KEY MANAGEMENT

Software-based key management adds multiple layers of policy and protection for any and all Hadoop encryption keys

SECURE SENSITIVE ARTIFACTS


Store, secure, and manage any sensitive, security-related object generated by Hadoop (i.e. keys, truststores, passphrases, and more) in a “virtual safe-deposit box”

QUICK AND EASY DEPLOYMENT

Software, available through Cloudera Navigator, can be deployed in hours rather than days or weeks

HSM INTEGRATION

Master encryption keys can be stored in environments with existing HSM-centric compliance and security policies



Data security is no longer a checkbox in the IT or operations departments, but is a top business priority for most enterprises. As compliance requirements like HIPAA and PCI-DSS continue to expand in scope, and as unstructured data—often sensitive information relating to customers and corporate IP—proliferates inside an organization, the requirements and restrictions on data and data access have come under increased scrutiny.

This need for heightened security comes at a time when organizations are finding real business value in platforms like Apache Hadoop for delivering more data across the enterprise. Cloudera Enterprise (powered by Hadoop) is a single, unified solution that lets you store and analyze all your data and metadata, while providing compliance-ready security and governance, and end-to-end system management.

This centralized form of data management presents tremendous opportunity for enterprises looking to unlock the value of their data, but it also represents a challenge; namely, how to protect the sensitive business data and associated artifacts in a manner that satisfies compliance, and meets internal and customer-driven security mandates.

Data Encryption Everywhere

Cloudera Enterprise is the only Hadoop platform to provide out-of-the-box encryption for both “data in motion,” between processes and systems, as well as “data-at-rest” as it persists on disk or other storage mediums. While the Heartbleed bug raised concerns about the vulnerability of data over the wire, the fact is that data-at-rest is far more susceptible to attack if left unprotected.

As part of Cloudera Navigator, Cloudera customers can now leverage industry standard AES-256 encryption for all HDFS files, HBase records, Hive metadata, and audit logs. The encryption runs at the filesystem level and is completely transparent to the applications reading and writing to disk; so performance overhead is minimal, and deployment is quick and painless. Organizations can gain additional performance improvements when running this type of encryption on Intel® hardware featuring the AES-NI cryptographic accelerator available on Xeon® and Core™ processors.

Bundled with at-rest encryption are unique process-based access controls. This method allows authorized Hadoop processes to access the encrypted data, while simultaneously preventing admins or super users, like root, from being able to run commands that grant them access to data they don't need to see.

Secure Key Management

Critical as at-rest encryption is to the Hadoop data protection landscape, it is equally important that the encryption keys are managed in a secure and compliant fashion. Key generation, storage, and access need to be carefully managed to avoid a breach and subsequent data loss. The same is true for other digital artifacts that control access to sensitive objects:

- _ Private keys for SSL certificates
- _ TrustStore files
- _ Passphrases
- _ Hadoop Java keystores

Previously, this type of robust key management system was beyond the scope of Hadoop. It was expensive and often required a hardware installation that complicated deployments and required configuration changes. But that's no longer the case for Cloudera customers.

Now, Cloudera Enterprise users can take advantage of software-based and policy-controlled key management that protects Navigator Encrypt keys and any other digital artifact that must be secured and controlled, per policy. In compliance with NIST requirements, the keys are always stored separate from the encrypted data and wrapped in multiple layers of cryptography.

Key management in Cloudera Enterprise functions like a "virtual safe-deposit box," supporting a variety of robust, configurable, and easy-to-implement policies governing access to the secure artifacts. Cloudera Navigator Key Trustee also integrates with a variety of hardware security modules (HSMs).

Data encryption and key management at multiple layers within Hadoop provide a critical layer of protection against potential threats by malicious actors on the network or in the data center. It's also a requirement for meeting key compliance initiatives and ensuring the integrity of your enterprise data.

OUT-OF-THE-BOX DATA PROTECTION FOR THE ENTERPRISE DATA HUB

At-Rest Encryption	<ul style="list-style-type: none">_ High-performance transparent data encryption for HDFS, Hive, HBase, and more_ Rapid deployment and configuration through Cloudera Navigator, requiring no changes to any Hadoop applications
Key Management	<ul style="list-style-type: none">_ Software-based key management with strong configurable management policies and lifecycle_ Any security-related object can be stored in a secure vault that allows for true separation of keys and objects from encrypted data
Access Controls	<ul style="list-style-type: none">_ Fine-grained access controls to data and metadata in Hadoop and role-based authorization through Apache Sentry_ Supports process-based access controls to prevent unauthorized users and systems from accessing sensitive data
Governance Solution	<ul style="list-style-type: none">_ A single interface for visibility into where data came from and how it's being used_ Comprehensive auditing, access controls, discovery and exploration, lineage, and lifecycle management

