Hortonworks®

Ensure PCI DSS compliance for your Hadoop® environment

A Hortonworks White Paper
October 2015

Contents

Overview

Big Data offers tremendous potential for merchants, banks and other commercial entities seeking to unlock new value from their customer and business data–but the Hadoop environment that powers these strategies can pose new challenges for compliance with the Payment Card Industry Data Security Standard (PCI DSS). Created by the Payment Card Industry Security Standards Council to increase controls around cardholder data and reduce the risk of fraud, the PCI standard spans twelve control objectives across six areas from network configuration to data storage as a security baseline for data storage, processing and transmission. As you develop your Big Data technology strategy, it's important to make sure that the architecture you build is designed to facilitate compliance with PCI. While PCI compliance applies to individual implementations or projects rather than technologies–which means that no product or vendor can be considered "PCI-compliant"–the right product will provide the capabilities you need to meet crucial PCI requirements.

Serving many of the world's largest retail, e-commerce and telecom companies, Hortonworks recognizes the critical importance of PCI compliance for our customers. The Hortonworks security architecture helps these businesses meet the data protection requirements of PCI through a holistic approach based on five pillars:

- Administration
- Authentication and perimeter security
- Authorization
- Audit
- Data protection

Each of these pillars complements the others as part of a unified, 100% open source platform that can in turn be extended through the additional security capabilities of our ecosystem partners.

By building essential capabilities for PCI compliance into the DNA of the Hortonworks Data Platform (HDP), Hortonworks helps merchants, banks and other businesses leverage the full value of Big Data while protecting their organization and its customers from risk.

Why PCI matters to your business

PCI compliance applies to all organizations or merchants, regardless of size or number of transactions, that accept, transmit or store any cardholder data. Businesses that fail to meet the requirements of this standard risk losing the ability to process these payments. Violations can also lead to fines as high as $500,000 per incident or $100,000 per month for non-compliance. Beyond the enforcement aspect of the standard, though, compliance with PCI DSS should be a central tenet of the business's strategy for security and data protection. The high-profile security breaches appearing regularly in the headlines provide vivid reminders of the damage companies can incur through inadequate or inconsistent security measures, from damaged customer relationships and corporate reputation to regulatory fines by government agencies.

Introduced in 2014, PCI v3.0 is the current standard for merchants, processors, payment service providers or anyone else storing or using payment card data. The standard covers all applications and systems in a merchant or a payment service provider, and encompasses the following 12 components:

PCI Data Security Standards – High Level Overview	
Build and maintain and secure network & systems	• Install & maintain a firewall configuration to protect card holder data • Do not use vendor-supplied default for system passwords and other security parameters
Protect cardholder data	• Protect stored cardholder data • Encrypt transmission of cardholder data across open, public networks
Maintain a vulnerability management program	• Protect all systems against malware and regularly update anti-virus software or programs • Develop and maintain secure systems and applications
Implement strong access control measures	• Restrict access to cardholder data by business need to know • Identify and authenticate access to system components • Restrict physical access to cardholder data
Regularly monitor and test networks	• Track and monitor all access to network resources & cardholder data • Regularly test security systems and processes
Maintain an information security policy	• Maintain a policy that addresses information security for all personnel

PCI is designed to protect both cardholder and sensitive authentication data relating to payment methods used by consumers.

- **Cardholder data** includes primary account number (PAN), cardholder name, expiration date and service code.
- **Sensitive authentication data** includes data stored in magnetic stripe or in chip (full track data), CVV numbers (the three- or four-digit number on the back or front of the card), and PINs or PIN-related information.

PCI compliance rules mandate different treatment for each of these types of data. Sensitive authentication data must not be stored in any part of the IT infrastructure. Card numbers (PAN) may be stored, but must be encrypted at all times, for example through tokenization or anonymization techniques.

	Type of data	Storage allowed	Encryption /Anonymization at REST required
Cardholder Data	Primary account number	Yes	Yes
	Cardholder Name	Yes	No
	Expiration Date	Yes	No
	Service code	Yes	No
Sensitive Authentication Data	Full track data	No	N/A (cannot store data)
	CVV numbers	No	N/A (cannot store data)
	PIN	No	N/A (cannot store data)

Building support for PCI compliance into your Hadoop environment

PCI compliance applies to all network, systems and applications with access to sensitive data such as payment information–including Hadoop clusters, which under the standard are treated the same way as any other ERP, EDW or other system that might be used to store and process such data. As a result, IT leaders must ensure that the Data Lake at the core of their Hadoop implementation meets the same high standards of security as any legacy data environment. The Hortonworks Data Platform (HDP) forms the core of the modern data architecture that allows enterprises to store and process massive amount of structured, semi-structured, and unstructured data. Trusted by more than 330 customers and proven in highly security-conscious environments, HDP provides robust security capabilities to help businesses meet PCI requirements.

Based a holistic approach to protection, the Hortonworks security framework revolves around five pillars: administration, authentication/ perimeter security, authorization, audit and data protection. Rather than applying protection as an afterthought, Hortonworks prevents gaps and inconsistencies through a bottom-up platform approach that makes it possible to enforce and manage security across the stack through a central point of administration and management built into the DNA of HDP. By implementing security at the platform level, Hortonworks ensures that security is consistently administered to all the applications across the stack, and makes the process of adding or retiring applications operating on top of Hadoop seamless.

The Hortonworks security architecture begins with Apache Ranger, Apache Knox and Kerberos. Ranger serves as the central interface for security administration. Users can create and update policies, which are then stored in a policy database. Ranger plugins consisting of lightweight Java programs are embedded within processes of each cluster component. These plugins pull

in policies from a central server and store them locally in a file. When a user request comes through the component, these plugins intercept the request and evaluate it against the security policy. Plugins also collect data from the user request and follow a separate thread to send this data back to the audit server.

This platform approach ensures that each of the five pillars of Hadoop security complement each other effectively to enable comprehensive protection.

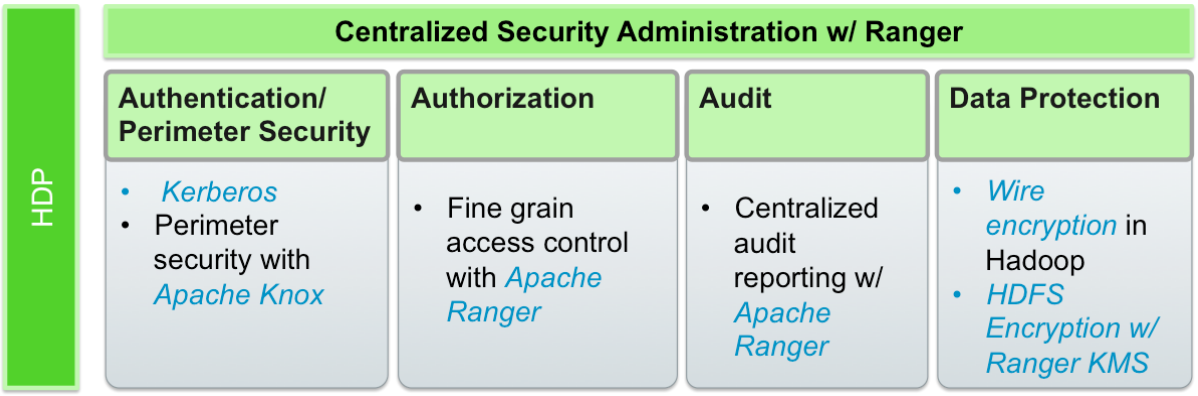

Figure 1: Comprehensive security in HDP

Administration

Ranger provides a single pane of glass to define, administer and manage security policies consistently across all the components of the Hadoop stack. Ranger enhances the productivity of security administrators and reduces potential errors by empowering them to define security policy once and apply it to all the applicable components across the Hadoop stack from a central location.

Authentication and perimeter security

Apache Knox Gateway ensures perimeter security for Hortonworks customers, providing a way for users to reliably identify themselves and then have that identity propagated throughout the Hadoop cluster to access resources such as files and directories, and to perform tasks such as running MapReduce jobs. Hortonworks uses Kerberos, an industry standard, to authenticate users and resources within Hadoop cluster. Hortonworks has also simplified Kerberos setup, configuration and maintenance through Ambari 2.0. With Knox, enterprises can confidently extend the Hadoop REST API to new users without Kerberos complexities, while also maintaining compliance with enterprise security policies. Knox provides a central gateway for Hadoop REST APIs that have varying degrees of authorization, authentication, SSL and SSO capabilities to enable a single access point for Hadoop.

Authorization

Ranger manages fine-grained access control through a rich user interface that ensures consistent policy administration across Hadoop data access components. Security administrators have the flexibility to define security policies for a database, table and column or a file, and administer permissions for specific LDAP based groups or individual users. Rules based on dynamic conditions such as time or geography can also be added to an existing policy rule. The Ranger authorization model is highly pluggable and can be easily extended to any data source using a service-based definition. Ranger works with standard authorization APIs in each Hadoop component and is able to enforce centrally administered policies for any method of accessing the data lake. The combination of Ranger's rich user interface with deep audit visibility makes it highly intuitive to use, enhancing productivity for security administrators.

Audit

Established by Hortonworks with Aetna, Merck, Target and SAS, the Data Governance Initiative (DGI) introduced a common approach to Hadoop data governance into the open source community. This initiative has since evolved into a new open source project called Apache Atlas, a set of core foundational governance services that enables enterprises to effectively and efficiently meet their compliance requirements within Hadoop and allows integration with the complete enterprise

data ecosystem.

Ranger also provides a centralized framework for collecting access audit history and easily reporting on this data, including the ability to filter data based on various parameters. Together with Apache Atlas, this makes it possible for users to gain a comprehensive view of data lineage and access audit, with an ability to query and filter audit based on data classification, users or groups, and other filters.

Data Protection

HDP adds a robust layer of security by making data unreadable in transit over the network or at rest on a disk. Hortonworks has recently introduced HDFS encryption for encrypting HDFS files, complemented with a Ranger-embedded open source Hadoop key management store (KMS). Ranger provides security administrators with the ability to manage keys and authorization policies for KMS. Hortonworks is also working extensively with its encryption partners to integrate HDFS encryption with enterprise-grade key management frameworks. With Hortonworks, our customers have the flexibility to leverage an open source key management system (KMS), or use enterprise wide KMS solutions provided by the partner ecosystem.

Encryption in HDFS, combined with KMS access policies maintained by Ranger, prevents rogue Linux or Hadoop administrators from accessing data and supports segregation of duties for both data access and encryption.

How Hortonworks helps you meet key PCI requirements

The Hortonworks security architecture can help customers with PCI requirements in these specific areas:

- The Data protection pillar helps you meet the requirement to protect cardholder data.
- The Authentication / Perimeter Security and Authorization pillars help you meet the requirement to implement strong access control measures.
- The Audit pillar helps you meet the requirement to regularly monitor and test networks
- The Administration pillar helps you define and apply policies effectively across each of these areas to support comprehensive PCI compliance.

Area	PCI Requirement	How Hortonworks can help?
Protect cardholder data	Protect and store credit card data	Hortonworks partners such as Voltage, Protegrity, Vormetric and Dataguise provide complementary capabilities such as tokenization and format-preserving encryption. Customers also have the option to use HDFS encryption with Ranger KMS to encrypt files in a specific directory
	Encrypt transmission of cardholder data across open, public networks	HDP supports wire encryption for all access protocols as well as Solr, Kafka and YARN, with dynamic attributes such as geo, time and data to drive security policy decisions.
Implement strong access control measures	Restrict access to cardholder data by business need-to-know	Apache Ranger provides fine-grained access control across HDFS, Hive, HBase, Storm, Knox, Solr, Kafka and Yarn. Apache Knox can be used to restrict access at the perimeter level.
	Assign a unique ID to each person with computer access	HDP supports implementation of Kerberos for authentication within a cluster. Apache Knox provides authentication for REST-based services and can be integrated with AD/LDAP or other authentication mechanisms.

Regularly monitor and test networks	Track and monitor all access to network resources and cardholder data	Apache Ranger provides auditing for users access across Hive, HDFS, HBase, Storm, Knox, Solr, Kafka and YARN.

A checklist for PCI compliance in Hadoop

As you begin the process of ensuring PCI compliance for your Hadoop environment, it can be helpful to think about the following questions to identify key areas of focus.

What are the PCI implications of your Hadoop use case? You should know what kinds of data Hadoop store in your organization and whether it might potentially contain payment card data.

Can you prevent raw payment card data from entering your Hadoop environment? Remember that PCI DSS strongly discourages companies from storing cardholder data such as name and account number, and requires this data to be encrypted or otherwise rendered unreadable in the event that it is stored. If the data imported into your Hadoop environment could contain payment card information, see if it can be obfuscated or tokenized first to ensure compliance with this requirement.

What steps have you already taken to protect data in Hadoop? These can include:

- Implementing Kerberos within clusters to protect identities
- Using Apache Ranger for fine-grained access control and auditing
- Using Apache Knox to protect API access
- Enabling wire encryption to protect data when it being transmitted
- Tokenizing card data when it is stored within Hadoop

Do you have monitoring in place to determine whether PCI-related data is entering Hadoop? Make sure you can scan the data landing in Hadoop regularly so you're not caught by surprise by improperly stored payment card data. Data discovery partners such as Dataguise can help in identifying sensitive data in Hadoop.

Is your Hadoop cluster an integrated part of your overall enterprise security architecture? Your Hadoop implementation should be protected through existing organizational security infrastructure such as firewall and network security measures, and governed by your company's information security policy. The physical servers running Hadoop should be physically protected in your data center.

Hortonworks security specialists can help in performing a security analysis of a Hadoop implementation and give recommendations on best practices for securing Hadoop and accelerate achieving PCI compliance.

The myth of the PCI-compliant Big Data product

One of the myths noted by the Payment Card Industry Security Standards Council is the existence of a "silver bullet"–a vendor or solution that companies can rely on for full PCI compliance. In reality, a vendor or product can't be PCI-compliant; only a project or deployment. Because every implementation is unique, a product used to achieve full compliance in one organization might leave another one far short of the mark. The goal is to find key capabilities that helps you deploy and maintain an implementation that achieves full compliance.

As the PCI Security Standards Council says, "Many vendors offer an array of software and services for PCI compliance. No single vendor or product, however, fully addresses all 12 requirements of PCI DSS." Rather than seeking a single silver bullet, you should make sure that the technologies you apply in each of these 12 areas provides capabilities to help you achieve compliance. If a single product can help you in more than one area, all the better.

Ultimately, there's no automatic way to achieve PCI compliance, and no substitute for the conscientious Implementation of carefully selected technologies.

Conclusion

To realize the full strategic benefits of Big Data without increasing the risk of compliance violations or data breaches, companies need to ensure that their Hadoop environments meet the strict requirements laid out in the PCI standard. The best way to accomplish this is through a holistic approach based on a platform that provides for all five essential pillars of Hadoop security–administration, authentication / perimeter security, authorization, audit and data protection. With HDP, Hortonworks provides robust capabilities to address key PCI requirements and facilitate the creation and maintenance of a fully compliant modern data architecture. More than 330 customers trust HDP to power their Big Data strategy, including some of the world's largest retail, e-commerce and telecom companies. To learn more about how Hortonworks can help you address key security and data protection challenges in your organization, we welcome you to download the Hortonworks white paper "Solving Hadoop Security" or speak with a Hortonworks representative at 1-855-8-HORTON.

About Hortonworks

Hortonworks develops, distributes and supports the only 100% open source Apache Hadoop data platform. Our team comprises the largest contingent of builders and architects within the Hadoop ecosystem who represent and lead the broader enterprise requirements within these communities. Hortonworks Data Platform deeply integrates with existing IT investments upon which enterprises can build and deploy Hadoop-based applications. Hortonworks has deep relationships with the key strategic data center partners that enable our customers to unlock the broadest opportunities from Hadoop. For more information, visit www.hortonworks.com.