HOW ZERO TRUST PRIVILEGE SOLVES FIVE HADOOP SECURITY RISKS
ABOUT CENTRIFY

Centrify is redefining the legacy approach to Privileged Access Management by delivering cloud-ready Zero Trust Privilege to secure modern enterprise use cases. Centrify Zero Trust Privilege helps customers grant least privilege access based on verifying who is requesting access, the context of the request, and the risk of the access environment. By implementing least privilege access, Centrify minimizes the attack surface, improves audit and compliance visibility, and reduces risk, complexity and costs for the modern, hybrid enterprise. Over half of the Fortune 100, the world’s largest financial institutions, intelligence agencies, and critical infrastructure companies, all trust Centrify to stop the leading cause of breaches — privileged credential abuse.

To learn more visit www.centrify.com.
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Executive Summary</td>
<td>1</td>
</tr>
<tr>
<td>With Big Data Comes Big Responsibility</td>
<td>2</td>
</tr>
<tr>
<td>Five Key Security Risks Associated with Hadoop</td>
<td>5</td>
</tr>
<tr>
<td>— and How to Avoid Them</td>
<td></td>
</tr>
<tr>
<td>The Centrify Solution</td>
<td>8</td>
</tr>
</tbody>
</table>
Executive Summary

The growth of big data over the last several years has been explosive and the rapid adoption of big data solutions will continue through the foreseeable future. This shouldn’t come as a surprise considering the value proposition of enabling organizations to analyze their most important data. Hadoop is the most significant technology in this big data revolution, and organizations adopting Hadoop stand to reap great rewards. Retail companies, for example, can benefit from this deep analysis of data to personalize every customer experience, anticipate outcomes, and assist in targeting the right prospect with the right offer at the right time.

The benefits of deploying Hadoop are significant, but with any infrastructure that stores an organization’s most valuable data, it’s essential that businesses are aware of the potential security risks and take adequate steps to address them. And the first step is to implement effective privileged access management.
With Big Data Comes Big Responsibility

The ever-increasing amount of personal customer information and critical corporate data means cyber attacks — both from within and outside the organization — can cause significantly more damage than ever before.

According to the Center for Strategic and International Studies (CSIS), the global cost of cyber crime has now reached as much as $600 billion — about 0.8 percent of the global GDP. More worrying than that figure may be the massive growth from 2014, when the same analysis showed the cost was only as much as $445 billion.

Whether on purpose or by accident, insiders are probably the weakest link in your organization, and the most likely culprit in any scenario that ends in data loss. In fact, many of today’s most high-profile breaches — while engineered by outsiders — are launched using the credentials of an insider. In the era of big data, securing information and the technologies that process it has never been more critical.

Hadoop Technology

As web data has exploded and grown beyond the ability of traditional systems to process it, Hadoop has emerged as the de facto standard for storing, processing and analyzing hundreds of terabytes of big data.

In its report on the Hadoop Market, Zion Market Research forecasts that the global Hadoop market will grow at a compound annual growth rate of almost 50% from 2017 to 2022. Global market revenue, estimated at $7.69 billion in 2016, is predicted to grow to $87.14 billion by the end of 2022.

"According to the 2018 Insider Threat Report, 90% of organizations feel vulnerable to insider threats and 53% even experienced insider attacks in the last 12 months."
Hadoop enables distributed parallel processing of huge amounts of data across inexpensive, industry-standard servers that also store the data, and can scale without limits. No data is too big for Hadoop. And in an environment where more data is created every day, Hadoop enables organizations to derive significant, meaningful value out of once unusable information.

The Hadoop environment is essentially a distributed database with a built-in computational cluster used to analyze data that came from many different sources, most often in search of patterns of behavior or risk. It can be used to tackle any number of big data business needs across a variety of industries including financial services, retail, national security, entertainment and many more.

The rapidly growing popularity of Hadoop speaks volumes to the value it delivers to organizations. But high-value data carries a high-risk potential. Today, big data is a prime target for security breaches because of the large volumes of valuable information it encompasses — including PII, PCI, HIPAA and other federally regulated data.

Hadoop Brings Significant Benefits, but Implementation Must Be Strategic

Hadoop environments have typically been set up by business analysts or developers as a mechanism to respond to questions posed by individual departments. Because it was originally designed for use on a private network by a limited number of designated users, security was not a primary consideration.

Developers have since enhanced the technology with components that allow it to be deployed in secure mode — incorporating Kerberos to authenticate from one node to the next, and encryption for data transport between nodes.

But enterprise-wide adoption still requires a strategic and adequately cautious approach:

- **Organizations must configure Hadoop in secure mode before it enters production.** By default, Hadoop runs in non-secure mode, and while businesses can set up an MIT Kerberos environment to ensure that each user and service is authenticated, implementing this system is typically a time-consuming, multi-step process that’s prone to error. Moreover, it creates a parallel identity infrastructure, redundant to most organizations’ Active Directory environments — which already provide Kerberos authentication capabilities.

- **Organizations must strictly control user access.** Granting the right privileged users access to the nodes in the cluster requires identity and access management. But many Hadoop admins are unfamiliar with centrally managing user accounts and their access to the cluster. And centralized access management is essential. Many organizations have hundreds or thousands of nodes inside multiple Hadoop clusters that, when managed manually, would require a user account to be set up on each individual node.

And that’s not a one-time exercise — the Hadoop ecosystem is constantly changing as new applications, interfaces and analysis engines become available, each with new and different methods of privileged user access. Management becomes even more complex when corporate and regulatory security requirements dictate that authentication and access controls must be consistently applied to each new interface.
How Zero Trust Privilege Solves Five Hadoop Security Risks

- **Organizations must control administrative privileges within the cluster.** In order to securely move to production, IT must centralize controls over privileged user access. Assigning IT admins specific privileges across the cluster allows local root accounts to be locked down. In order to protect the jobs submitted by privileged users and client applications, as well as the data they access, security staff must have verifiable control over IT staff access to and privileges for managing clusters. But such control and visibility is a challenge and businesses remain hard-pressed to find a simple way to manage these privileges across the distributed Hadoop cluster.

**Securing Hadoop With Your Current Identity Management Infrastructure**

Security-adept organizations are leveraging their existing Active Directory infrastructure, skill sets and management processes to address a number of key Hadoop security concerns. The integration of Hadoop clusters and their supporting applications into Active Directory can provide centralized identity management that further enables the adoption of a true, cross-platform privilege management and auditing solution across Hadoop clusters, nodes and services. Leveraging existing Active Directory accounts to log in, secures Hadoop environments and assists in proving compliance in a repeatable, scalable and sustainable manner—and without deploying and managing a new identity infrastructure.

Once cluster nodes are integrated with Active Directory, only the addition of new service accounts is required for automated authentication from one node to the next. Active Directory authentication also allows for single sign-on for Hadoop administrators and end users, which helps to reduce Hadoop deployment and management costs and increase worker productivity.

An identity management infrastructure makes it easy and cost effective for Hadoop organizations to:

- **Deploy Hadoop clusters in Secure Mode** by leveraging Active Directory's Kerberos capabilities for secure mode and automating the configuration of Hadoop service accounts.
- **Simplify and standardize identity and access management**, leveraging Active Directory group-based access controls for Hadoop cluster access management.
- **Automate manual, error-prone processes** used for on-going management of secure user and application access to Hadoop clusters and between Hadoop services.
- **Allow only authorized administrators** to manage the Hadoop cluster.
- **Remove anonymity** in the processing of Hadoop jobs by attributing privileged actions to an individual.
Five Key Security Risks Associated with Hadoop — and How to Avoid Them

There are a few specific security risks associated with Hadoop that can be addressed by integrating the cluster into Active Directory. In an all-Windows Hadoop environment, integration into Active Directory is a relatively simple process, and can deliver effective access management. In more common Linux Hadoop environments, an additional solution can be employed to achieve integration between the Linux cluster — or any non-Windows cluster — and Active Directory.

But access is only half of the equation. A solution should not only provide access management across Windows, Linux and UNIX servers, it should also provide comprehensive, centralized identity management that includes privileged access management and auditing capabilities, which can be extended across the entire Hadoop environment. Because they leverage your existing Active Directory infrastructure, skill sets and other investments, these solutions have been shown to deliver cost savings as well. The result is significantly improved management and security, as well as the ability to avoid the following five risks, typical in Hadoop environments:

1. Yet another application identity silo

   In the rush to realize immediate value, many big data teams are building complex and disparate identity management infrastructures (identity silos) resulting in increased security exposure and risk, not to mention increased implementation and operations expenditures around big data and identity management deployments. Identity silos in Hadoop environments:
   
   - Lack enterprise integrated user and admin access control
   - Lack visibility over Hadoop user activity and client applications that submit data jobs
   - Are typically managed by non-identity professionals
   - Increase the risk of failed compliance audits
   - Place additional pressure on IT resources to manage a rapidly growing number of Hadoop clusters and data nodes with new infrastructure and more identities

   A centralized, cross-platform identity management infrastructure removes the need for identity silos across Hadoop clusters, nodes and services. By simply implementing a solution that leverages Active Directory, IT can grant access to Hadoop clusters using existing identities and group memberships, versus creating new identities for users across every Hadoop cluster.

2. Increased internal and external threat potential

   Without centralized control over who can access Hadoop clusters (including data nodes), how and when these users can access the cluster and what they can do with it, organizations are leaving the back door unlocked for both internal and external attacks to occur. One disgruntled employee that gains access to the company’s Hadoop ecosystem can retrieve vast amounts of valuable data.

   By integrating the Hadoop environment into the existing identity management infrastructure, user authentication is simplified and access is tightly enforced.
How Zero Trust Privilege Solves Five Hadoop Security Risks

For access to jobs and data, both user identity and group membership may be required, and even those with credentials are limited to access the results from their own data and MapReduce jobs. Authenticated users are granted the ability to access both Name Nodes as well as any Data Node they’re required to access, but one user cannot view the results of another’s job.

The auditing capabilities that come with a comprehensive identity management solution can further ensure that all user activity is tied to a specific individual, making it easy to identify who did what across Hadoop clusters, nodes and services.

3. Runaway admin privileges

Giving administrators more privileges than they require to do their job is risky in any situation. But giving them more privileges than they require in Hadoop environments can be disastrous given the vast amount of highly valuable data available.

It’s a known fact that, whether intentional or by accident, insider threats account for roughly 50% of the breaches and data loss businesses incur today. And it’s therefore essential that once inside the cluster, administrators are granted a more narrow set of access and privilege than the local root account would allow. Administrator access should follow a policy of least privilege — strictly limited to the specific actions and commands necessary to execute their job functions and responsibilities — not full access to the cluster and all its data.

Once Hadoop clusters, nodes, services, applications and users are integrated into Active Directory, a centralized, full-function identity management solution should also provide an easy way to granularly grant privileges to the appropriate user based on their role or job function.

The result is an environment where users gain privileged access to specific functions, data and nodes, without gaining access to the root account or the entire data lake within the Hadoop cluster.

At the same time, auditors gain much more visibility into the actions of IT staff, developers and users through session monitoring, which attributes all privileged actions to a clearly identified, individual Active Directory account.

Because there are times when users require additional access privileges for a single project, a temporary role, or to participate in a program outside their normal job scope, privilege elevation can be granted to specific Hadoop users or roles for specified time frames. Users as well as administrators should use privilege elevation to extend explicit privilege rights as required, and again, should never be given the root user account.
User session activity monitoring can be integrated with existing SIEM log monitoring solutions to provide greater accountability with full visibility of all user activity.

4. Decreased mitigation and remediation capabilities

Without visibility into what’s happening across the Hadoop cluster, it’s virtually impossible to identify, mitigate and remediate potential security issues. It’s therefore imperative that organizations ensure security and accountability by auditing user activity across the cluster — and particularly those clusters that contain sensitive business data or personally identifiable information.

A full-function identity management solution should provide user session monitoring that includes the detailed capture of all user activity. This allows businesses to establish accountability by recording which users accessed individual systems inside the Hadoop environment, what commands they executed, with what privilege, and to identify the exact changes they made to key files and configurations, if any.

Enterprises can then pinpoint suspicious activity through monitoring, troubleshoot system failures by replaying actions for root-cause analysis, and report on historical user activity. Effective auditing can detect fraud, inappropriate entitlements, unauthorized access attempts and a number of other anomalies that may be missed with normal event log monitoring. Out-of-the-box and custom reports can be created to prove to auditors which systems each specific user can access, and which users have access to each specific system.

Session activity auditing can help organizations to reduce the severity of both inside and outside breaches while also helping to ascertain the specific actions associated with the attack, discover accountability and prevent future attacks.

5. Diminished regulatory compliance

Improperly deployed and managed Hadoop clusters can increase the risk of failed SOX, GLBA, PCI DSS, HIPAA, FISMA, NERC, MAS compliance audits.

The key components of many compliance regulations include tightly enforced access controls with least privilege access policy enforcement and user session auditing.

So it’s essential that organizations enable repeatable, enterprise-wide auditing and compliance that includes big data solutions across all nodes.

Leveraging Active Directory through an identity management solution can help to streamline regulatory compliance across the Hadoop ecosystem through:

- Tightly enforced access controls
- Centrally managed least privileged security policies
- User activity auditing associated with Active Directory individual user accounts

A large financial services company originally implemented Hadoop to gain customer intelligence. Impressive results led to an extended Hadoop ecosystem, and today the company has thousands of nodes across multiple clusters, with plans for additional growth. The company uses Active Directory to control access to its Hadoop system and remain in compliance with MAS regulations.
The Centrify Solution

Centrify Zero Trust Privilege for Hadoop leverages Active Directory to provide comprehensive identity and access management, privilege management and session auditing for Hadoop environments.

Benefits

Secure and simplify Hadoop environments without the hassle of deploying and managing new identity infrastructure. Centrify makes it simple to run Hadoop in secure mode by leveraging your existing identity management infrastructure — Active Directory. Minimize costs by eliminating the need for parallel solutions — that may not be enterprise ready — and leveraging existing skill sets.

Increase productivity and security via single sign-on for IT admins and Hadoop users. By extending the power of Active Directory's Kerberos and LDAP capabilities to your Hadoop clusters, Centrify delivers authentication for Hadoop administrators and end users.

Secure machine-to-machine communication across the nodes in your cluster. Centrify automates Hadoop service account management. Centrify not only secures user identity but also system and service account identity.

Mitigate threats resulting from identity-related risks and address compliance and audit requirements. Hadoop environments store your organization’s most important data. Centrify controls access, manages privileges, audits activity and associates everything back to an individual, thereby making your data more secure. Centrify delivers cost-effective compliance with reporting on who has access and who did what across Hadoop clusters, nodes and services.
Our mission is to stop the leading cause of breaches – privileged access abuse. Centrify empowers our customers with a cloud-ready Zero Trust Privilege approach to secure access to infrastructure, DevOps, cloud, containers, Big Data and other modern enterprise use cases. To learn more, visit www.centrify.com.

Centrify is a registered trademark of Centrify Corporation. Other trademarks mentioned herein are the property of their respective owners.

©2019 Centrify Corporation. All Rights Reserved.