KuppingerCole Report

EXECUTIVE VIEW

by Mike Small | January 2016

Centrify for Big Data

Centrify Server Suite integrates Hadoop and NoSQL clusters into Microsoft Active Directory for user authentication, authorization and auditing.

by **Mike Small**
mike.small@kuppingercole.com
January 2016

Content

Related Research Documents

Advisory Note: From Big Data to Smart Information - 70750

Advisory Note: Big Data Security, Governance, Stewardship - 71017

Executive View: Big Data and Information Stewardship - 70744

Leadership Brief: Privileged Account Management Considerations - 72016

Leadership Compass: Privilege Management - 71100

Executive View: Centrify Identity Service - 71186

Executive View: Centrify Server Suite - 70886

1 Introduction

There is now an enormous quantity of data available in a wide variety of forms and more is continuously being generated. Collectively known as "Big Data", this includes the vast amount of data that organizations have accumulated internally as well as that from external sources. These include social media and publicly available data from government databases as well as other data shared between organizations.

Big Data technologies were invented to store and process this vast amount of data into useable "Smart" Information. The most commonly mentioned tool is Hadoop[1] which was developed by Yahoo and released as an open source tool written in Java on top of Apache. This provides a way of searching large data sets in parallel using commodity computing hardware.

Big Data requires more flexibility than is provided by conventional relational databases. This has led to the growth of the so called "NoSQL" type of database. These databases use looser consistency models than traditional relational databases in order to achieve horizontal scaling and higher availability. They are often optimized for appending and retrieval operations.

What is common across these technologies is that their initial aims focused on data processing capabilities rather than security and compliance. One particular concern has been lack of control over identity and access especially in the area of administration. This was fine when the application of the tools was confined to experimental or small scale usage. Now that they are being widely deployed for commercial application this is no longer satisfactory.

Organizations using Big Data need to remain compliant with a wide range of laws and regulations. Big Data can be misused through abuse of privilege; curiosity may lead to unauthorized access and information may be deliberately leaked. Mistakes can also lead to disclosure of sensitive information and incorrect analysis can lead to incorrect or inappropriate conclusions.

As described in KuppingerCole Advisory Note: "Big Data Security, Governance, Stewardship" - 71017 an information centric approach to big data is needed to ensure:

- Availability: individuals are able to access the Big Data and Smart Information they need to perform their business functions when and where they need it, and without delay.
- Integrity: individuals are only able to manipulate Big Data (create, change or delete) in ways that are authorized.
- Confidentiality: Big Data and Smart Information can only be accessed by authorized individuals and these are not able to pass data to which they have legitimate access to other individuals who are not authorized.

This can only be achieved with appropriate identity and access management for the big data technologies.

[1] http://developer.yahoo.com/hadoop/

2 Centrify for Big Data

Centrify is a US based Identity Management software vendor that was founded in 2004. Centrify has achieved recognition for its identity and access management solutions for web and cloud-based applications, as well as management for Mac and mobile devices and their apps. The company is Venture Capital funded and has raised significant funding from a number of leading investment companies. The company as of today has more than 5,000 customers.

Centrify is best known for their capability of integrating UNIX and Linux account management into Microsoft Active Directory, but also supports integration of Mac OS X. This still is at the core of their Centrify Server Suite. However, the overall portfolio of Centrify has grown, adding the Centrify Identity Service for secure access to Cloud applications as well as Centrify Privilege Service for shared account password management and secure remote access to servers, while the features of the Centrify Server Suite and its various editions have now been extended to include Big Data clusters.

The Centrify Server Suite is a comprehensive identity and access management system. It now enables the centralized management - based around Active Directory - of identity and access to a range of platforms used for the storage, processing and analysis of Big Data.

2.1 Centrify Server Suite for Big Data

Centrify Server Suite integrates the nodes in Hadoop and NoSQL clusters providing user authentication and leveraging centralized access controls defined within Microsoft Active Directory.

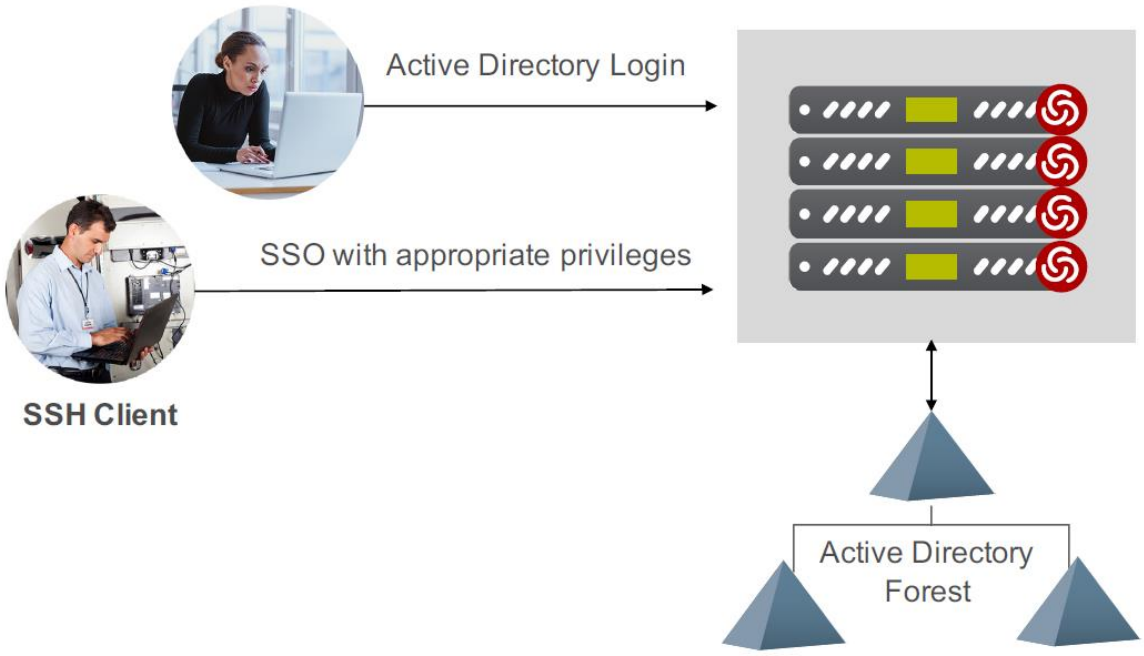

Figure 1: Centrify for Big Data Overview

2.1.1 Centrify and Hadoop

Hadoop is a cluster of computers built using off the shelf hardware that can process very large sets of data in parallel. These computers comprise a large number of worker data nodes and a smaller number of master nodes that control the distribution and processing of data. The software environment is written in Java. This architecture leads to information security and compliance challenges around the administration of the cluster, controlling access to the data, the cluster and its components, as well as auditing.

Centrify Server Suite is certified for a number of Hadoop environments including: Cloudera Enterprise, Hortonworks Data Platform and MapR Enterprise Edition.

2.1.2 Centrify and NoSQL

NoSQL databases provide a mechanism for storing and processing data that is not modelled in relational tables. Their architecture is optimized to increase the speed of certain kinds of query and analysis to be above that possible using the relational model. In particular NoSQL databases support horizontal scaling where a query can be processed in parallel across a number of computers. Hence NoSQL databases are often deployed across large clusters of computers and this raises the same, previously described, challenges around administration and controlling access to the data, the cluster and its components, as well as auditing.

Centrify Server Suite is certified for NoSQL databases provided by Cloudera, DataStax and MongoDB

2.2 Centrify Functionality for Big Data

Centrify for Big Data provides a range of common functionality and benefits for both Hadoop and NoSQL clusters used for processing Big Data.

2.2.1 Microsoft Active Directory Integration

Centrify Server Suite integrates the nodes in both Hadoop and NoSQL clusters into Microsoft Active Directory providing centralized user authentication, privilege management and auditing.

For Hadoop Clusters Centrify Server Suite automates the configuration of Hadoop in secure mode centrally, leveraging Active Directory. Centrify generates service accounts within Active Directory and distributes service account credentials across all nodes. This central management of service accounts and auto-configuration of Kerberos to work with Active Directory simplifies running Hadoop in secure mode.

The cluster nodes can be auto-joined to Active Directory using tools that automate the process of installing Centrify software and integrating servers with Active Directory, even complex environments with multiple domains and cross-forest trusts. The Centrify solution provides both administrators and big data analysts with single-sign on using their standard Active Directory credentials. It eliminates the need to create multiple identities for users that creates a “one user, one identity” framework that strengthens security and reduces IT administration.

For NoSQL clusters Centrify Server Suite provides single sign-on to IT administrators and users using Centrify-enabled PuTTY (SSH client) with their standard Active Directory credentials. NoSQL typically uses LDAP to communicate to Active Directory; Centrify’s LDAP Proxy integrated with Centrify’s Active

Directory agent provides access to Active Directory environments including those with multiple domains or cross-forest relationships.

2.2.2 Role Based Privilege Management

Centrify Server Suite helps to extend privileged user management to include Big Data clusters. This exploits Centrify's patented Zone technology to configure roles and grant users the appropriate privileges and access to resources. Administrators always log in as themselves — not as root or Local Admin — then only elevate privileges when needed.

2.2.3 Session Auditing

Centrify Server Suite maintains a central record of all administrator and user session activity. This provides the information required by auditors to demonstrate compliance with regulations. It also enables an organization to ensure user accountability through correlating activity across the clusters.

2.2.4 VPN-Less Remote Access

Centrify Identity Service enables cloud, mobile and on-premises app single sign-on (SSO) for users, and a simplified identity infrastructure for IT. The App+ edition of Identity Service adds the App Gateway, which enables secure remote access and single sign-on to on-premises web apps, including those used by outsourced developers and administrators for Big Data, without need to install and maintain VPNs. Centrify Privilege Service extends this secure remote access to administrative users who need SSH or RDP access to Linux or Windows consoles.

2.2.5 Lockdown of administrative accounts

Centrify Privilege Service provides a cloud-based solution to secure, periodically rotate and checkout local administrator accounts such as root or local admin. These account passwords can either be stored within the encrypted storage of the cloud service or optionally stored on-premises within a SafeNet Secure Store as the primary repository. This service also provides administrators with full control over account password checkout or to provide remote secure sessions using these accounts eliminating the administrator's need to check-out the password.

3 Strengths and Challenges

The emerging technologies for the storage, processing and analysis of Big Data pose many challenges in the areas of security and compliance; not least in the area of managing, controlling and auditing identity and access. Centrify Server Suite provides a potential solution for these problems to organizations that use Microsoft Active Directory as their central user repository – which applies to most organizations. It integrates the management and authentication of users to Big Data appliances, like Hadoop and NoSQL database clusters, into Microsoft Active Directory. It provides fine-grained, role-based authorization and monitoring of the use of these platforms integrated with other platforms deployed within the organization.

The security and compliance challenges of Big Data extend beyond those of managing identity and access.

These include aspects such as control over the provenance of the data, the ownership of external data, and the classification of the data as well as the results of analysis. Identity and access management are important components of a solution to these challenges but are not, in themselves, a comprehensive solution. An ideal security and compliance solution for Big Data would include identity and access management capabilities integrated with information centric security capabilities covering all aspects. There is currently no such solution on the market.

Until this kind of comprehensive solution exists, organizations that are looking to secure the storage, processing and analysis of Big Data should seriously consider the Centrify Identity Platform.

Strengths	Challenges
• Authentication and fine grained authorization of users to Big Data Clusters. • Integrated management of identity and access privileges to multiple platforms including Big Data around Active Directory. • Role based management of entitlements. • Comprehensive control over privileged user access. • Support for session monitoring and auditing, including capturing of meta-data. • Secure remote access to Big Data without the need for VPN.	• While this is an excellent solution for organizations already using Active Directory as their main identity store it may be less attractive to those using other solutions for this. • Big Data has security and compliance challenges beyond control over identity and access. Ideally the solution should be part of a complete solution.

4 Copyright

The Future of Information Security – Today

KuppingerCole supports IT professionals with outstanding expertise in defining IT strategies and in relevant decision making processes. As a leading analyst company KuppingerCole provides first-hand vendor-neutral information. Our services allow you to feel comfortable and secure in taking decisions essential to your business.

KuppingerCole, founded in 2004, is a global Analyst Company headquartered in Europe focusing on Information Security and Identity and Access Management (IAM). KuppingerCole stands for expertise, thought leadership, outstanding practical relevance, and a vendor-neutral view on the information security market segments, covering all relevant aspects like: Identity and Access Management (IAM), Governance & Auditing Tools, Cloud and Virtualization Security, Information Protection, Mobile as well as Software Security, System and Network Security, Security Monitoring, Analytics & Reporting, Governance, and Organization & Policies.

For further information, please contact **clients@kuppingercole.com**

Kuppinger Cole Ltd.
Sonnenberger Straße 16
65193 Wiesbaden | Germany

Phone +49 (211) 23 70 77 – 0
Fax +49 (211) 23 70 77 – 11
www.kuppingercole.com