Intel Hadoop UCS Lab

Key Security Recommendations

By Brad Antoniewicz

Prepared May 13, 2015

Proprietary and confidential

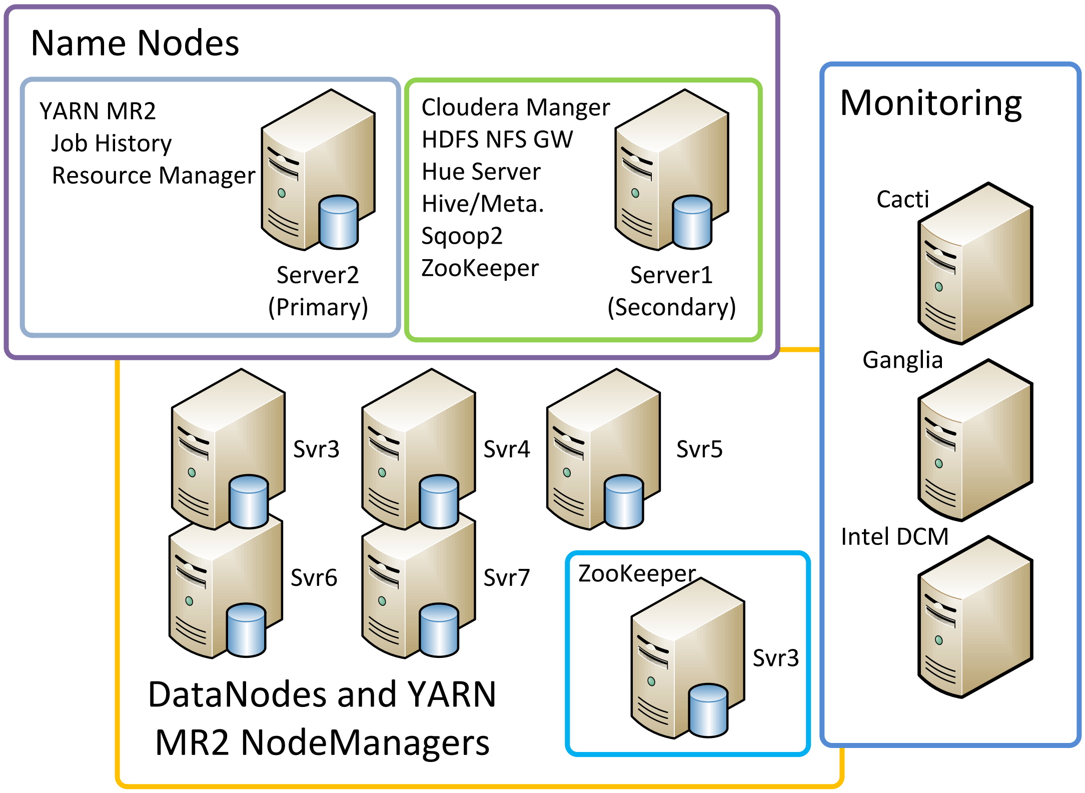
PROBONO REPORT

Overview

Security with the Hadoop environment is commonly controlled at the perimeter, regulating who can communicate and submit jobs to the nodes of the cluster. While this is a good first level defense, the relatively open design of the Hadoop ecosystem makes compromising the entire system somewhat trivial once this single layer is defeated. To help better guide security within Hadoop environments, we'll look at a typical deployment and make recommendations which will lessen the impact in the case of a perimeter breach.

Environment

Typical to most environments, the servers within ours serve multiple roles:



Each server is built with the default installation of CentOS 6.5 and exists within a Cisco UCS environment supported by two Cisco UCS 6300 Fabric Interconnects. The cluster is managed by Cloudera Manager CDH 5.

Attack Surface

When considering security, perhaps the most urgent step in a deployment is to limit the available ports and services exposed by a system. Each open port offers a potential avenue to system compromise for an attacker. This is can be a complex task in a Hadoop environment given the variety of roles any particular node may have. To illustrate this point, let's look at what's available on the first three servers in our lab environment:

Host	Protocol	Port	Purpose
Server 1	TCP	22	SSH
	TCP	111	RPC
	TCP	2049	NFS
	TCP	2181	ZooKeeper Client Port
	TCP	4181	ZooKeeper Election Port
	TCP	4242	HDFS NFS Gateway
	TCP	5678	Cloudera Reports Manger Server Port
	TCP	7180	Cloudera Manager Admin Console
	TCP	7182	Cloudera Agent Connect Port
	TCP	7184	Cloudera Event Publish Port
	TCP	7185	Cloudera Event Query Port
	TCP	7432	Postgres
	TCP	8083	Cloudera Reports Manager Debug WebUI
	TCP	8084	Cloudera Event Server Debug WebUI
	TCP	8086	Cloudera Service Monitor Debug WebUI
	TCP	8087	Cloudera Activity Monitor Debug WebUI
	TCP	8091	Cloudera Host Monitor Debug WebUI
	TCP	8649	Ganglia
	TCP	8888	Hue Server
	TCP	9000	Cloudera Agent
	TCP	9010	ZooKeeper JMX Remote Port
	TCP	9083	Hive Metastore Server Port
	TCP	9994	Cloudera Host Monitor Nozzle Port
	TCP	9995	Cloudera Host Monitor Listen Port
	TCP	9996	Cloudera Service Monitor Nozzle Port
	TCP	9997	Cloudera Service Monitor Listen Port
	TCP	9998	Cloudera Activity Monitor Nozzle Port
	TCP	9999	Cloudera Activity Monitor Listen Port
	TCP	10000	Hive Server2 Port
	TCP	10101	Cloudera Alert Publisher Listen Port
	TCP	12000	Sqoop2 HTTP Port
	TCP	40164	ZooKeeper
	TCP	50090	Secondary NameNode Web UI Port
Server 2	TCP	22	SSH

	TCP	111	RPC
	TCP	8040	Node Manager Localizer Port
	TCP	8041	NodeManager IPC Port
	TCP	8042	NodeManager Web UI Port
	TCP	8649	Ganglia
	TCP	9000	Cloudera Agent
	TCP	13562	MR2 Shuffle Port
	TCP	50010	DataNode Transceiver Port
	TCP	50070	NameNode WebUI Port
	TCP	50075	DataNode HTTP Web UI Port
Server 3	TCP	22	SSH
	TCP	111	RPC
	TCP	8040	Node Manager Localizer Port
	TCP	8041	NodeManager IPC Port
	TCP	8042	NodeManager Web UI Port
	TCP	8649	Ganglia
	TCP	9000	Cloudera Agent
	TCP	13562	MR2 Shuffle Port
	TCP	50010	DataNode Transceiver Port
	TCP	50020	DataNode Protocol Port
	TCP	50075	DataNode HTTP Web UI Port

As you can see we have a considerable amount of potential entry points for an attacker, a vulnerability in any one of these services could lead to local system compromise and ultimately cluster data.

In the few sections we'll outline the major points an attacker would target given a route to the target environment. These sections are mostly the more appealing to an attacker and shouldn't be considered an exhaustive list.

Web UIs

Nearly each component within the Hadoop ecosystem has a web interface. The power of these web interfaces vary greatly, with some being just information dumps while others provide query or even job submission capabilities. By default, most of these web interfaces lack login functionality so anyone with network level access to the Hadoop environment can access them. These UIs offer a variety of risks such as:

- Unprotected log information may be used in more sophisticated attacks
- The web UI itself may be vulnerable to attack (XSS, Injection, etc...)
- The web UI may provide privileged functionality such a job submission, HDFS navigation, configuration changes, etc…

RPC Services

The services provided within a Hadoop solution commonly rely on one another and need to communicate autonomously. They do this through the use of Remote Procedure Call (RPC) services, commonly in the form of a WebAPI. Be default, these services may or may not require authentication, which means an attacker with network level access to the system may be able to directly query, submit jobs, or even access the data stored on the sever.

Linux Services and Local Processes

It's common to enable various more common Linux services and local processes to support the set up and usage of the cluster. Each of these services and local processes also provide a potential point of entry to a remote attacker or means of escalation to a local attacker.

Data in Transit and at Rest

Hadoop leverages data in a variety of contexts, some of that data may be the actual data stored on the HDFS, while other data might be the configuration files for the services, or telemetry data sent to/from monitoring system. Regardless of what the data is and in what context it's utilized, it most likely needs to be protected as failure to do so may provide an advantage to an attacker.

Key Points in Securing Hadoop

Next we'll look at the major steps in securing a Hadoop environment.

Segmentation

This section is mostly to emphasize perimeter defense mentioned in the overview of this paper as the bare-minimum. Even within an organizationally segmented network, the Hadoop environment should be further segmented by placing it behind a firewall with conservative port restrictions to limit traffic to/from it.

Lab Environment Assessment

The Lab Environment appeared to be appropriately segmented, however it was outside of the scope to review the network architecture of the environment so only a cursory review was completed.

General OS Hardening

Each of the systems within the cluster and those responsible for the management and monitoring of the cluster should undergo general OS hardening. This can be easily achieved by using a pre-hardened VM as the base for all systems. General OS hardening includes checks that cover:

1. Updates and Patching
2. File system permissions
3. User and group configuration
4. Password rotation, complexity, etc...
5. Secure boot
6. Local intrusion detection
7. Process hardening
8. Service restrictions
9. Remote access policies

The Center for Internet Security (http://www.cisecurity.org) offers OS hardening guides called benchmarks and tools to perform automated analysis of existing systems (e.g. CIS-CAT). Nessus, McAfee Vulnerability Manager, and various other commercial and open source products exist to routinely inspect the configuration of the system

Lab Environment Assessment

- **Missing Critical Patches** - The DataNodes and supporting systems were running an outdated but supported CentOS 6.5 distribution. Systems were vulnerable to many software issues that could be easily remediated through a standard system update. A traditional vulnerability scan was not conducted due to the segmentation of the network.
- **Inadequate System Hardening Configuration –** Leveraging the CIS-CAT utility, a benchmark test was run and revealed a number of inadequate system configuration settings ranging from missing banners and insufficient logging to improper password complexity settings. A copy of the output from the CIS-CAT utility is included with this document as a supplement.

General Hadoop Security Configuration

The Hadoop project offers a "Secure Mode" (http://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-common/SecureMode.html) configuration that guides cluster administrators on best practice for the overall cluster. The security options within Hadoop are categorized as:

1. User and Daemon Authentication
http://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-common/SecureMode.html#Authentication

2. Service Level Authorization
http://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-common/ServiceLevelAuth.html

3. Web Console Authentication
http://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-common/HttpAuthentication.html

4. Data Encryption/Confidentiality
http://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-common/SecureMode.html#Data_confidentiality

5. Map Reduce Encrypted Shuffle
http://hadoop.apache.org/docs/current/hadoop-mapreduce-client/hadoop-mapreduce-client-core/EncryptedShuffle.html

6. HDFS Permissions
http://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-hdfs/HdfsPermissionsGuide.html

The "Secure Mode" guide and those associated with the security categories above provide detailed configuration steps and should be followed carefully, testing each component individually to ensure any service impact is identified.

Kerberos Deployment

The ecosystem relies heavily on Kerberos to authenticate users and services as they interact with the various components of the ecosystem. This requires a Kerberos Domain Controller (KDC). Many organizations may have an existing KDC and may be tempted to integrate it directly with the Hadoop environment, however that should be avoided. A dedicated Hadoop KDC should be deployed for service accounts and the Enterprise KDC should have a one-way trust set up with the Hadoop KDC for user accounts. This allows user accounts to remain consistent, and reduce the management for service accounts by cluster administrators.

Management Platforms

One benefit of utilizing a management platform such as Cloudera Manager is that some components of "Secure Mode" may likely be implemented by default. For instance the following items from the "Secure Mode" guide were found to be implemented:

1. **User Accounts for Hadoop Daemons** – By default Cloudera deployed individual service accounts for each daemon. Accounts cloudera-scm, hue, hive, hdfs, impala, and sqoop2 were all created and assigned to a hadoop group.
2. **Local System File Permissions** – Cloudera also sets local file system permissions appropriately for each user account daemon identified above.

Lab Environment Assessment

- **Secure Mode Disabled –** Hadoop's secure mode was not enabled completely on the lab environment and thus the Hadoop ecosystem would be subject to a myriad of attacks if the perimeter was breached.

Cloudera Security Configuration

The Cloudera Manager achieves control over the Hadoop Ecosystem through the use of a number services:

- Manager WebUI
- Cloudera Management Service Activity Monitor
- Cloudera Management Service Alert Publisher
- Cloudera Management Service Event Server
- Cloudera Management Service Host Monitor
- Cloudera Management Service Reports Manager
- Cloudera Management Service Service Monitor

Each of these services maintains its own configuration which adds an additional layer of effort in order to secure.

- **Management Security:** The Cloudera Manager itself should be appropriately secured as it is the central point for each of the services and serves as the Hub for the cluster.
- **Debug WebUI:** By default, each of these services are configured to provide a Debug WebUI on the server they're running. This may reveal sensitive information to an attacker or provide the ability to perform actions via the service.
- **Configuration Change Alerts:** In order to ensure unauthorized changes are detected, it is recommended that Cloudera be configured to issue configuration change alerts whenever a change is made.
- **Heap Dumps:** If a particular Service experiences an OutOfMemoryError, Cloudera offers the option for a heap dump to be performed and its results stored to disk. Sensitive data may be contained in such a file and so they should be disabled by default. Note: By default, most services do not have heap dumps enabled and have directory settings that limit risk.
- **Alert Reporting:** Security-related alerts, such as Configuration Changes should be configured to send an email or SNMP trap so that appropriate teams are notified.
- **Cloudera Usage Collection:** Usage data may be sent to Cloudera, it is unclear what exactly this data is. To protect the confidentiality of the environment, this should be disabled.

Cloudera Manager WebUI

- **Configuration Section:** Administration

Configuration Option	Property	Default	Recommended
Item: Manager Security **Sub Category:** Advanced	Enable Debugging of API	False	False
Item: Manager Security **Sub Category:** Advanced	Enabled Client Config Cleanup	True	True

Item: Manager Security **Sub Category:** Security	Session Timeout	30 mins	30 mins
Item: Manager Security **Sub Category:** Security	Allow 'Remember Me' option	True	False
Item: Manager Security **Sub Category:** Security	HTTP Referer Check	True	True
Item: Manager Security **Sub Category:** Security	Use TLS Encryption for Admin Console (Requires Keystore Configuration)	False	True
Item: Manager Security **Sub Category:** Security	Use TLS Encryption for Agents	False	True
Item: Manager Security **Sub Category:** Security	Use TLS Authentication of Agents to Server	False	True
Item: Manager Security **Sub Category:** Security	Path to TLS KeyStore File		<Path>
Item: Manager Security **Sub Category:** Security	Keystore Password		<Password>
Item: Manager Security **Sub Category:** Security	Path to TLS Truststore File		<Path>
Item: Manager Security **Sub Category:** Security	Truststore Password		<Password>
Item: Manager Security **Sub Category:** Security	Show Stacktraces on Error Pages	True	False
Item: Cloudera Usage Data Collection **Sub Category:** Other	Allow Usage Data Collection	True	False
Item: Cloudera Usage Data Collection **Sub Category:** Support	Send Diagnostic Data to Cloudera Automatically	True	False
Item: Cloudera Usage Data Collection **Sub Category:** Support	Use HTTPS to Upload Diagnostic Data	False	True
Item: Manager Security **Sub Category:** External Authentication	Authentication Backend Order*	Database Only	External than Database
Item: Manager Security **Sub Category:** External Authentication	External Authentication Type* (As long as some external authentication type is configured in the following options)	Active Directory	Active Directory
Item: Manager Security **Sub Category:** Parcels	Create Users and Groups, and Apply File Permissions for Parcels	True	True

*=Organizationally unique item.

Service-Wide

- **Configuration Section:** Clusters -> Cloudera Management Service
- **Category:** Service-Wide

Configuration Option	Property	Default	Recommended
Item: Configuration Change Alerts **Sub Category:** Monitoring	Enable Configuration Change Alerts	False	True

Cloudera Management Service Activity Monitor

- **Configuration Section:** Clusters -> Cloudera Management Service
- **Category:** Activity Monitor Default Group

Configuration Option	Property	Default	Recommended
Item: Debug WebUI **Sub Category:** Ports and Addresses	Activity Monitor Web UI Port	8087	-1
Item: Configuration Change Alerts **Sub Category:** Monitoring	Enable Configuration Change Alerts	False	True
Item: Heap Dumps **Sub Category:** Advanced	Dump Heap When Out of Memory	False	False

Cloudera Management Service Alert Publisher

- **Configuration Section:** Clusters -> Cloudera Management Service
- **Category:** Alert Publisher Default Group

Configuration Option	Property	Default	Recommended
Item: Configuration Change Alerts **Sub Category:** Monitoring	Enable Configuration Change Alerts	False	True
Item: Heap Dumps **Sub Category:** Advanced	Dump Heap When Out of Memory	False	False

Alert Reporting

Either email alerts or SNMP alerts as solutions that address this item, it is up to the organization to make the appropriate decision as to which one best aligns with their environment.

SMTP Alerting

Configuration Option	Property	Default	Recommended
Item: Alert Reporting **Sub Category:** None, Alert Publisher Default Group	Alerts: Enable Email Alerts	True	True
Item: Alert Reporting **Sub Category:** None, Alert Publisher Default Group	Alerts: Mail Server Protocol*	SMTP	SMTPS
Item: Alert Reporting **Sub Category:** None, Alert Publisher Default Group	Alerts: Mail Server Hostname	localhost	<Internal Mail Server>
Item: Alert Reporting **Sub Category:** Ports and Addresses	Alerts: Mail Server TCP Port		<Internal Mail Server Port>
Item: Alert Reporting **Sub Category:** None, Alert	Alerts: Mail Server Username		<Mail Server Username>

Publisher Default Group	(Only use with SMTPS)		
Item: Alert Reporting **Sub Category:** None, Alert Publisher Default Group	Alerts: Mail Server Password (Only use with SMTPS)		<Mail Server Password>
Item: Alert Reporting **Sub Category:** None, Alert Publisher Default Group	Alerts: Mail From Address	noreply@ localhost	noreply@ clouderamanger
Item: Alert Reporting **Sub Category:** None, Alert Publisher Default Group	Alerts: Mail To Address	root@ locahost	<Internal distribution list>

*=Organizationally unique item, recommended value is most secure.

SNMP Alerting

Configuration Option	Property	Default	Recommended
Item: Alert Reporting **Sub Category:** SNMP	SNMP NMS Hostname		<Internal SNMP Server>
Item: Alert Reporting **Sub Category:** SNMP	SNMP Server Port	162	<Internal SNMP Server Port>
Item: Alert Reporting **Sub Category:** SNMP	SNMP Security Level*	SNMPv2	AuthNoPriv
Item: Alert Reporting **Sub Category:** SNMP	SNMP Authentication Protocol	SHA	SHA
Item: Alert Reporting	SNMP Server Engine Id*		<SNMP Server Engine Id>
Sub Category: SNMP	SNMP Security Username*		<SNMP Username>
Item: Alert Reporting	SNMP Authentication Protocol Passphrase*		<SNMP Passphrase>

*=Organizationally unique item, recommended value is most secure.

Cloudera Management Service Event Server

- **Configuration Section:** Clusters -> Cloudera Management Service
- **Category:** Event Server Default Group

Configuration Option	Property	Default	Recommended
Item: Debug WebUI **Sub Category:** Ports and Addresses	Event Server Web UI Port	8084	-1
Item: Configuration Change Alerts **Sub Category:** Monitoring	Enable Configuration Change Alerts	False	True
Item: Heap Dumps **Sub Category:** Advanced	Dump Heap When Out of Memory	False	False

Cloudera Management Service Host Monitor

- **Configuration Section:** Clusters -> Cloudera Management Service
- **Category:** Host Monitor Default Group

Configuration Option	Property	Default	Recommended
Item: Debug WebUI **Sub Category:** Ports and Addresses	Host Monitor Web UI Port	8091	-1
Item: Configuration Change Alerts **Sub Category:** Monitoring	Enable Configuration Change Alerts	False	True
Item: Heap Dumps **Sub Category:** Advanced	Dump Heap When Out of Memory	False	False

Navigator Audit Server

- **Configuration Section:** Clusters -> Cloudera Management Service
- **Category:** Navigator Audit Server Default Group

Configuration Option	Property	Default	Recommended
Item: Debug WebUI **Sub Category:** Ports and Addresses	Navigate Audit Server Web UI Port	8089	-1
Item: Configuration Change Alerts **Sub Category:** Monitoring	Enable Configuration Change Alerts	False	True
Item: Heap Dumps **Sub Category:** Advanced	Dump Heap When Out of Memory	False	False

Navigator Metadata Server

- **Configuration Section:** Clusters -> Cloudera Management Service
- **Category:** Navigator Metadata Server Default Group

Configuration Option	Property	Default	Recommended
Item: Configuration Change Alerts **Sub Category:** Monitoring	Enable Configuration Change Alerts	False	True
Item: Heap Dumps **Sub Category:** Advanced	Dump Heap When Out of Memory	False	False
Item: Cloudera Usage Data Collection **Sub Category:** Advanced	Allow Usage Data Collection	True	False

Cloudera Management Service Reports Manager

- **Configuration Section:** Clusters -> Cloudera Management Service
- **Category:** Reports Manager Default Group

Configuration Option	Property	Default	Recommended

Item: Debug WebUI **Sub Category:** Ports and Addresses	Reports Manager Web UI Port	8083	-1
Item: Configuration Change Alerts **Sub Category:** Monitoring	Enable Configuration Change Alerts	False	True
Item: Heap Dumps **Sub Category:** Advanced	Dump Heap When Out of Memory	False	False

Cloudera Management Service Service Monitor

- **Configuration Section:** Clusters -> Cloudera Management Service
- **Category:** Service Monitor Default Group

Configuration Option	Property	Default	Recommended
Item: Debug WebUI **Sub Category:** Ports and Addresses	Service Monitor Web UI Port	8086	-1
Item: Configuration Change Alerts **Sub Category:** Monitoring	Enable Configuration Change Alerts	False	True
Item: Heap Dumps **Sub Category:** Advanced	Dump Heap When Out of Memory	False	False

Lab Environment Assessment

- **Server to Agent Security:** When user inquiries about any particular node, Cloudera Manager reaches out to that system over the applicable TCP port and requests the data related to the user's inquiry. By default, that transaction happens over the clear, allowing anyone with visibility into the environment to see the request. Authentication is performed by the "token" HTTP Cookie, which did not appear to change during the assessment. It was possible to sniff this token, and reuse it to issue arbitrary requests to the server.
- **Diagnostic Data:** It was surprising to see that by default, the manager sends Usage and Diagnostic Data to Cloudera on regular basis. Furthermore, the default setting as to whether or not to use HTTPS is set to not use it.

HDFS Security Configuration

The Hadoop Distributed File System (HDFS) is the primary location for cluster data and thus is extremely important to secure. The tradition of openness in the Hadoop Ecosystem can seem to be counterproductive to security, especially when it comes to data security. Outside of the permissions detailed within the Hadoop "Secure Mode" guide, the major components of securing HDFS relates to limiting the supplemental services that may facilitate unauthorized access to file system. There are various other mechanisms available to support encrypting data stored on HDFS while it is at rest, however this is outside of the scope of this evaluation.

NFS Gateway

The Network File System (NFS) Gateway allows remote systems to mount the HDFS store over the network. The default suggested configuration of the gateway leaves the share totally open, permitting access to anyone who can reach the service.

Configuration Option	Property	Default	Recommended
Cloudera HDFS Configuration • Category: NFS Gateway Default Group • Sub Category: None/NFS Gateway Default Group File: hdfs-site.xml	dfs.nfs.exports.allowed.hosts	* rw	<Specific Hosts>

Lab Environment Assessment

- **Unrestricted Access:** The lax permissions on the NFS gateway and the lack of "Secure Mode" on the NameNode Manager (TCP Port 50070) allowed access to the HDFS without authentication.

Additional Notes

This section contains various observations about the environment that could not be expanded on within the timeframe of the assessment.

- **Ganglia** - Ganglia's WebUI is unauthenticated and does not require TLS. Additionally the server queries the gmond daemon running on each node using a clear text protocol and the server itself is written in C, which makes it a candidate for fuzzing to identify memory corruption vulnerabilities

- **Hue** - Hue was left unconfigured. When connecting to the WebUI, Hue required that a username and password be set. Tester set it to root:password

- **Oozie/Hive Metastore/Zookeeper** - Given their functionality, Oozie, the Hive Metastore and Zookeeper were attractive targets for an attacker, however the Oozie deployment was not functioning so it was not possible to test and the others could not be reviewed in the timeframe of the assessment.

- **User Accounts** - All job submission via CLI and interaction with the nodes of the cluster was using the root user, this was not highlighted in the above sections but is an obvious security concern. Additionally the user rkypriot existed with a configured password. It was unclear if this was a service account of some sort or a legitimate user account. If it's a service account, the fact it has a password and can be logged into is concerning.

Administrative Contacts

Name	Title	Email	Phone Numbers
Foundstone			
Brad Antoniewicz	R&D	Brad_antoniewicz@mcafee.com	347-801-5864
Intel			
Patrick Schots	EMEA Technical Account Manager	patrick.schots@intel.com	patrick.schots@intel.com