What is Data Security?

Data Security for Hadoop allows you to administer a singular policy for authentication of users, authorize data access, protect data at rest and in motion, and collect a record of interactions with that data.

It allows you to coordinate enforcement of this policy across the entire Hadoop stack.
What is Security?

5 areas of security focus

Administration
Centrally management & consistent security

Authentication
Authenticate users and systems

Authorization
Provision access to data

Audit
Maintain a record of data access

Data Protection
Protect data at rest and in motion
Why Security?

5 areas of security focus

**Administration**
Centrally management & consistent security

**Authentication**
Authenticate users and systems

**Authorization**
Provision access to data

**Audit**
Maintain a record of data access

**Data Protection**
Protect data at rest and in motion

Security needs are changing
- YARN unlocks the data lake
- Multi-tenant: Multiple applications for data access
- Changing and complex compliance environment
- ETL of non-sensitive data can yield sensitive data

Summer 2014
65% of clusters host multiple workloads

Fall 2013
Largely silo’d deployments with single workload clusters
HDP delivers comprehensive security for Hadoop

COMPREHENSIVE SECURITY
Meet all security requirements across authentication, authorization, audit & data protection

CENTRAL ADMINISTRATION
Provide one location for administering security policies and for viewing and managing audit across the platform

CONSISTENT INTEGRATION
Integrate with other security and identity management systems, for compliance with IT policies
Start of 2014: Hadoop Data Security Capabilities

Requirements: Authenticate, authorize, provide auditability of data access and protect data at rest and in motion

<table>
<thead>
<tr>
<th>Administration</th>
<th>Did not exist</th>
</tr>
</thead>
<tbody>
<tr>
<td>Centrally management &amp; consistent security</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Authentication</th>
<th>Kerberos &amp; Apache Knox</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authenticate users and systems</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Authorization</th>
<th>Fragmented across Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provision access to data</td>
<td></td>
</tr>
<tr>
<td>• Hive: ATZ-NG</td>
<td></td>
</tr>
<tr>
<td>• HDFS: ACL’s</td>
<td></td>
</tr>
<tr>
<td>• Sentry (interactive SQL &amp; Search)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Audit</th>
<th>Fragmented across Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maintain a record of data access</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data Protection</th>
<th>3rd party add-ons, Integration w/Apache Falcon &amp; OS capabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protect data at rest and in motion</td>
<td></td>
</tr>
</tbody>
</table>

Start of 2014: State of Hadoop Security

• Largely disjoint patchwork
• A few projects address spot needs
• No coverage across all workloads
• No central administration and enforcement
May 2014: Hortonworks Acquires XA Secure

Requirements: Authenticate, authorize, provide auditability of data access and protect data at rest and in motion

<table>
<thead>
<tr>
<th>Administration</th>
<th>Authenticate</th>
<th>Authorization</th>
<th>Audit</th>
<th>Data Protection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Centrally management &amp; consistent security</td>
<td>Authenticate users and systems</td>
<td>Provision access to data</td>
<td>Maintain a record of data access</td>
<td>Protect data at rest and in motion</td>
</tr>
</tbody>
</table>

Hortonworks acquired XA Secure

Accelerates delivery against the enterprise requirements for central security administration and enforcement across all Hadoop workloads from, batch, interactive SQL and real-time

Founded in 2013, XASecure provides an enterprise ready, cross-platform, security layer built from the ground up for Hadoop, providing centralized capabilities around data security, authorization, auditing and overall governance.
Current State of Security in HDP

HDP provides central administration and coordinated enforcement of security policy across the entire Hadoop ecosystem of projects.

<table>
<thead>
<tr>
<th>Administration</th>
<th>Central administration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Centrally management &amp; consistent security</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Authentication</th>
<th>Kerberos &amp; Apache Knox</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authenticate users and systems</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Authorization</th>
<th>Authorization for HDFS, Hive, HBase, etc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provision access to data</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Audit</th>
<th>Compliance controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maintain a record of data access</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data Protection</th>
<th>3rd party add-ons, Integration w/Apache Falcon &amp; OS capabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protect data at rest and in motion</td>
<td></td>
</tr>
</tbody>
</table>

With additional XA Secure features, HDP is a leader in Hadoop Security

NEW FEATURES

- Centralized security policy enforcement
- Granular access control across HDFS, Hive, and HBase
- Universal audit tracking
- Compliance conformance controls
Central Security Administration

HDP Advanced Security

- Delivers a ‘single pane of glass’ for the security administrator
- Centralizes administration of security policy
- Ensures consistent coverage across the entire Hadoop stack
- Apache Argus

All delivered in Open Source under the governance of the ASF in Fall 2014
For more than 20 years, Kerberos has been the de-facto standard for strong authentication ... no other option exists.

What does Kerberos Do?

• Establishes identity for clients, hosts and services
• Prevents impersonation/passwords are never sent over the wire
• Integrates w/ enterprise identity mgmt tools such as LDAP & Active Directory
• More granular auditing of data access/job execution

The design and implementation of Kerberos security in native Apache Hadoop was delivered by Hortonworker, Owen O’Malley in 2010
Perimeter Security with Apache Knox

Incubated and led by Hortonworks, Apache Knox provides a simple and open framework for Hadoop perimeter security.

Single, simple point of access for a cluster
- Single Hadoop access point
- REST API hierarchy
- Consolidated API calls
- Multi-cluster support

Central controls ensure consistency across one or more clusters
- Eliminates SSH “edge node”
- Central API management
- Central audit control
- Simple Service level Authorization

Integrated with existing systems to simplify identity maintenance
- SSO Integration – Siteminder and OAM*
- LDAP & AD integration
Authorization and Audit

Authorization
Fine grain access control
• HDFS – Folder, File
• Hive – Database, Table, Column
• HBase – Table, Column Family, Column

Audit
Extensive user access auditing in HDFS, Hive and HBase
• IP Address
• Resource type/ resource
• Timestamp
• Access granted or denied
Data Protection

HDP allows you to apply data protection policy at three different layers across the Hadoop stack

<table>
<thead>
<tr>
<th>Layer</th>
<th>What?</th>
<th>How?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storage</td>
<td>Encrypt data while it is at rest</td>
<td>Partners, OS level encrypt, Custom Code</td>
</tr>
<tr>
<td>Transmission</td>
<td>Encrypt data as it moves</td>
<td>Supported in HDP 2.1</td>
</tr>
<tr>
<td>Upon Access</td>
<td>Apply restrictions when accessed</td>
<td>Partner, Custom Code</td>
</tr>
</tbody>
</table>
Thank You!

Eric Mizell – Director, Solution Engineering