Apache Knox - Hadoop Security Swiss Knife

Krishna Pandey, CISSP , CCSP

Staff Software Engineer (Platform Security)

Larry McCay

Senior Manager (Security Architecture) &

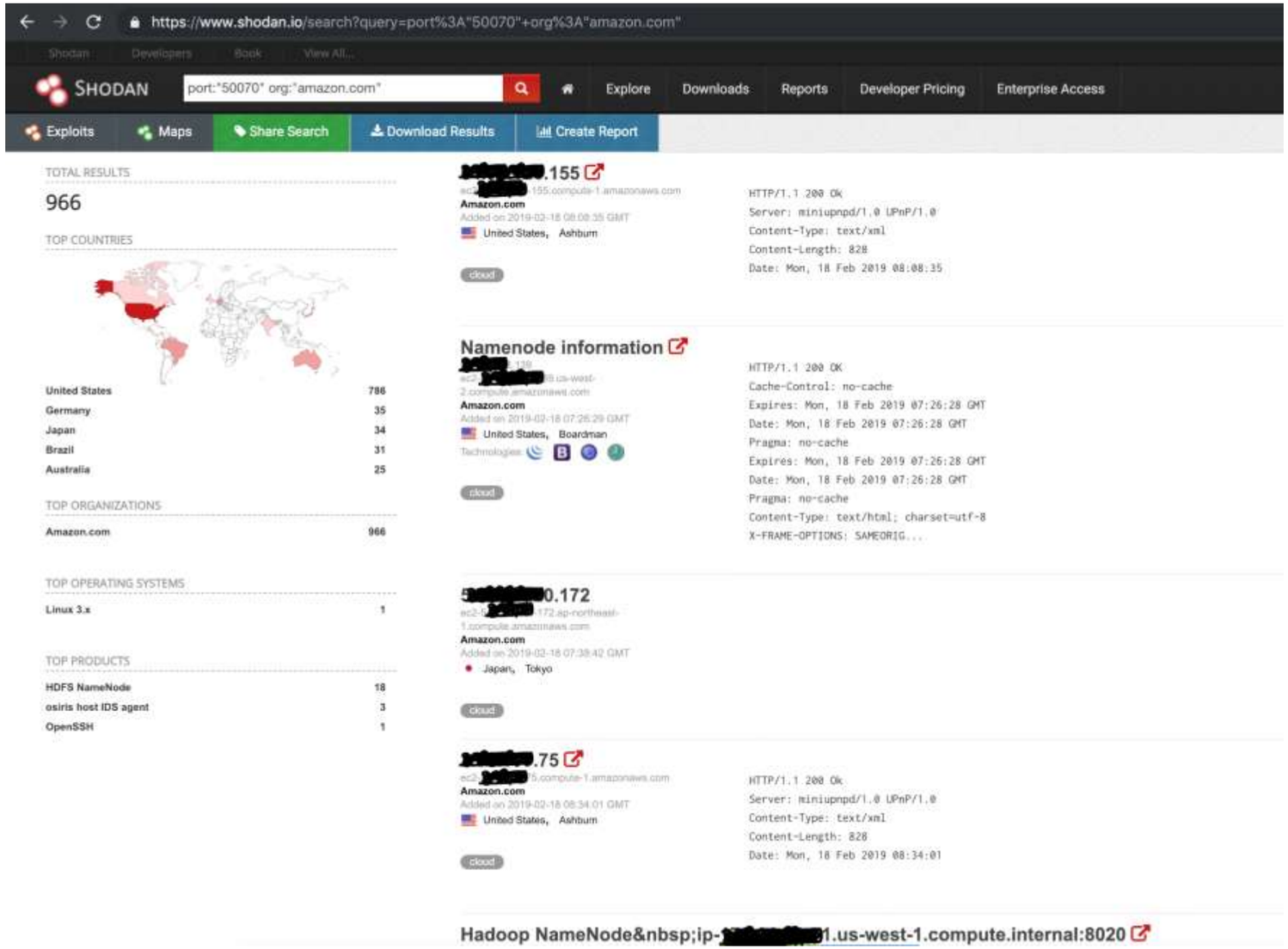
Knox PMC chair, Ranger, Metron and Hadoop committer

- Over 53,000 Incidents [1]
- 2,216 Confirmed Data Breaches [2]x
- Formjacking – 48,000 unique websites on average/month [3]
- Crypto-jacking [reduced by 52% in 2018] [4]
- Penalties for non-compliance – GDPR
- More CVEs to patch – CVE-2019-5736, CVE-2018-8009, …
- Meltdown, Spectre, Foreshadow, …, Spoiler

Source: [1] [2] https://www.verizonenterprise.com/verizon-insights-lab/dbir/
[3] [4]: https://www.symantec.com/content/dam/symantec/docs/reports/istr-24-executive-summary-en-apj.pdf

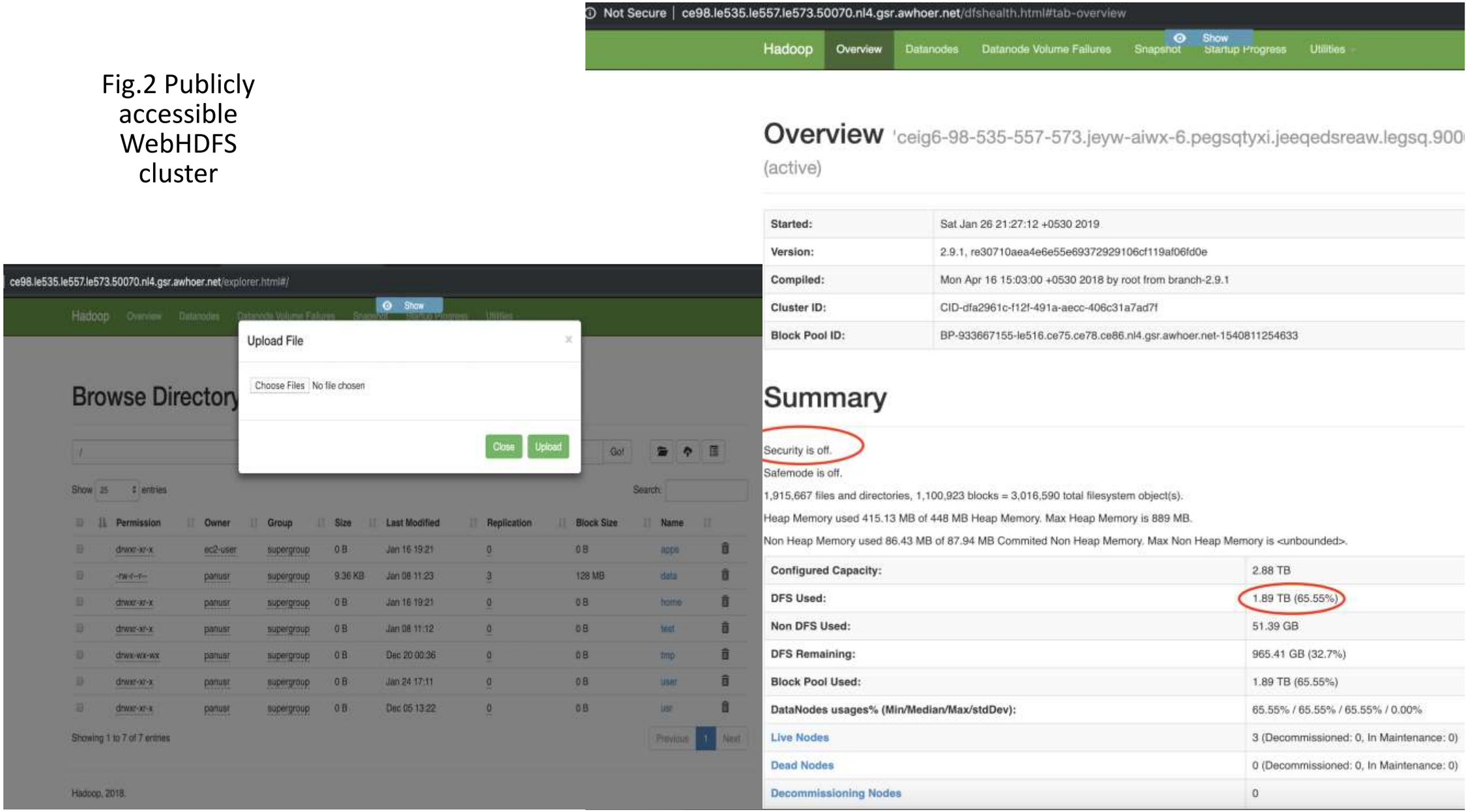
HORTONWORKS

What This Means to Big Data Landscape?



HORTONWORKS

Fig.2 Publicly accessible WebHDFS cluster

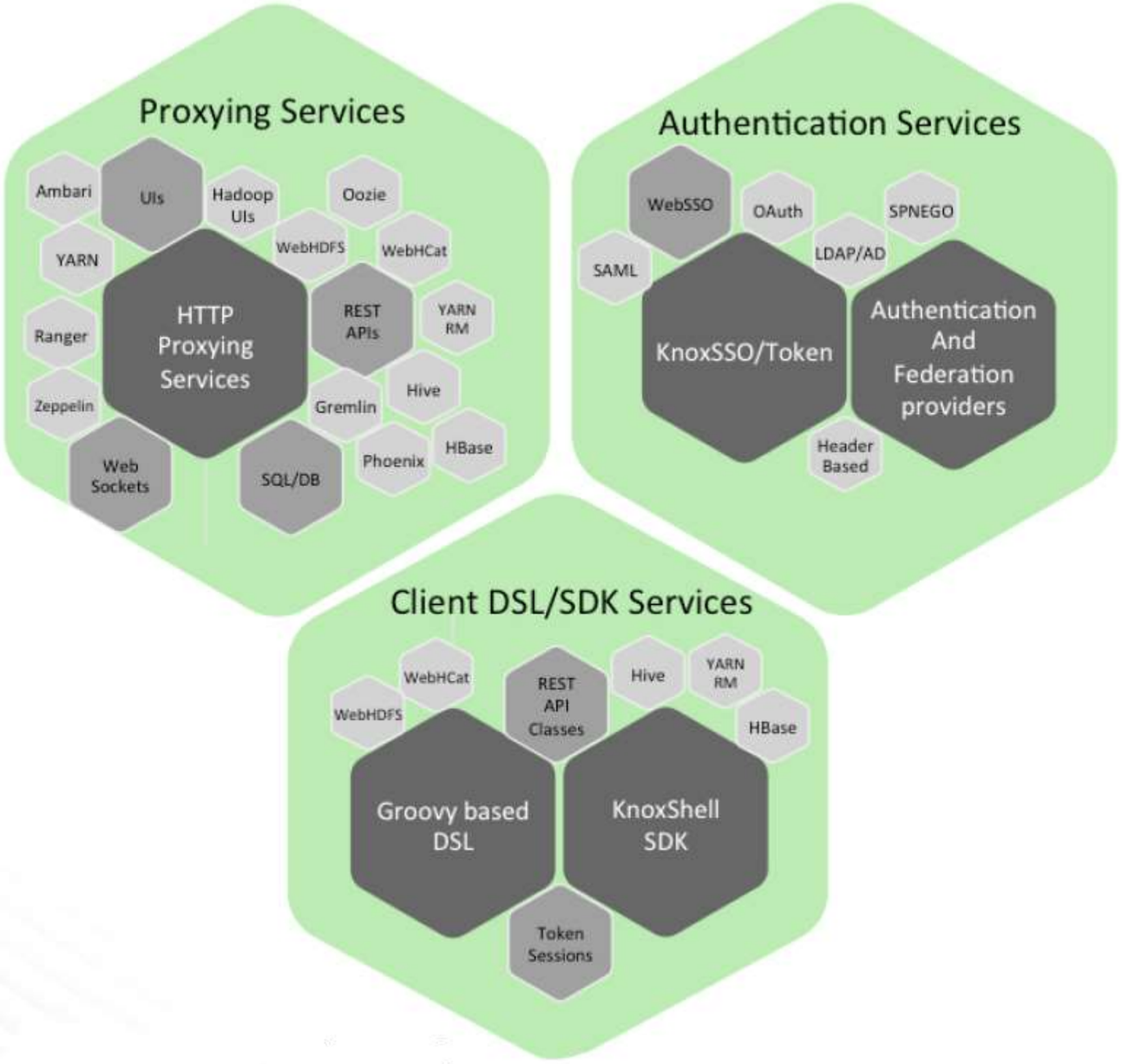


Apache Knox

- Knox Gateway is a reverse proxy for the Hadoop Ecosystem
- A trusted proxy within Hadoop platform deployments
- Proxies HTTP APIs and UIs
- Encapsulates and minimizes the burden of Kerberos on HTTP clients
- Highly flexible with pluggable provider chains

HORTONWORKS

Apache Knox – What is it?



3 Primary Sets of Services from Knox:

- Proxying Services
  - Protected Access to Platform Resources
  - Pluggable Provider Pipeline
- Authentication Services
  - KnoxSSO
  - KnoxToken
- Client Services
  - SDK Client Classes
  - Groovy based DSL for Scripting
  - Knox Token Sessions
  - Credential Collectors

HORTONWORKS

Why Knox?

Enhanced Security

- Proxy to abstract network details
- TLS Termination for non-SSL services

Centralized Control

- Auditing
- Service-level authorization
- Knox Admin UI
- Service Discovery and Topology Generation Framework

Simplified Access

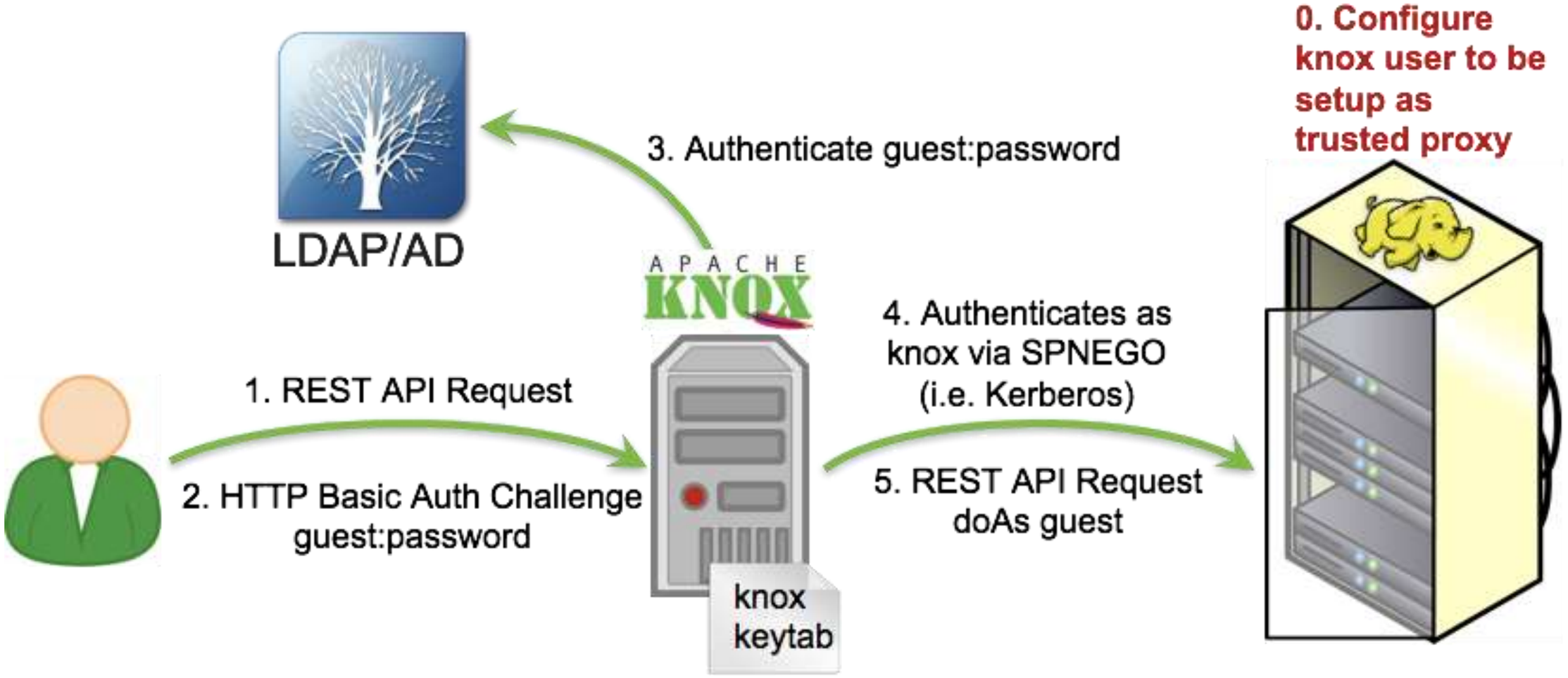
- Kerberos encapsulation
- Extends API reach
- Single access point
- Single SSL certificate
- Multi-cluster support

Enterprise Integration

- LDAP/AD integration
- Support for SAMLv2, OAuth and related IdPs: eg. Okta
- SSO integration

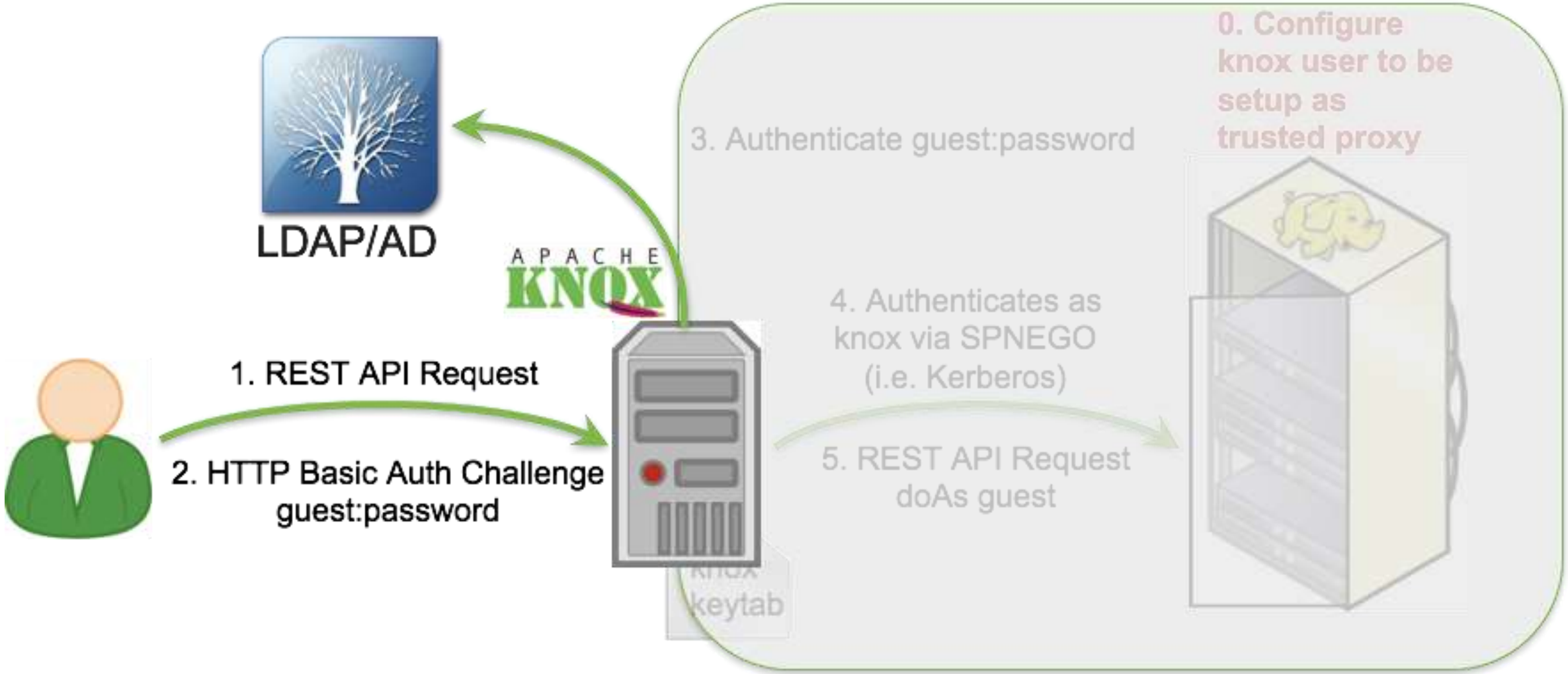
HORTONWORKS

Knox - Hadoop Trusted Proxy & Authn Service



HORTONWORKS

Encapsulation of Kerberos



HORTONWORKS

Apache Knox Providers

Web App Security	Protections against common web application vulnerabilities
Authentication / Federation	Authenticate a user's credentials or federate an external authentication event
Authorization	Verify that the authenticated user has access to requested resources
Audit	Detailed accounting of resource access

HORTONWORKS

Web Application Security

HORTONWORKS

Web App Security Provider

- Knox is a Web API (REST) and UI Gateway for Hadoop. The fact that REST interactions are HTTP based means that they are vulnerable to a number of web application security vulnerabilities.
  - CSRF
  - CORS
  - X-Frame-Options
  - X-Content-Type-Options
  - HTTP Strict Transport Security
  - X-XSS-Protection

HORTONWORKS

Authentication Mechanisms

HORTONWORKS

Apache Knox - Supported Authentication Mechanisms

- PAM based Authentication
  - provide the ability to use the configuration provided by existing PAM modules
  - remove limitations with the KnoxLdapRealm
    - Supported for nested OUs and nested groups
    - Faster lookups
    - Support more complex LDAP queries
    - Reduce load on the LDAP/AD server (caching by SSSD)

- HadoopAuth Authentication Provider
  - allows clients to using the strong authentication and SSO capabilities of Kerberos.
  - integrates the use of the Apache Hadoop module for SPNEGO and delegation token based authentication

HORTONWORKS

Supported Authentication Mechanisms (Contd.)

- Header based Preauthenticated SSO Provider
  - This provider was designed for use with identity solutions such as those provided by CA's SiteMinder and IBM's Tivoli Access Manager.
  - **CAUTION:** *The use of this provider requires that proper network security and identity provider configuration and deployment does not allow requests directly to the Knox gateway. Otherwise, this provider will leave the gateway exposed to identity spoofing.*

- Pac4j Provider
  - used as a federation provider to support the OAuth, CAS, SAML and OpenID Connect protocols.
  - It must be used for SSO, in association with the KnoxSSO service and optionally with the SSOCookieProvider for access to REST APIs.

Supported Authentication Mechanisms (Contd.)

- SSO Cookie Provider
  - a typical SP initiated websso mechanism that sets a cookie to be presented by browsers to participating applications and cryptographically verified.
  - enables the federation of the authentication event that occurred through KnoxSSO
- KnoxSSO Service
  - Provides WebSSO capabilities to the Hadoop cluster*
- KnoxToken Service
  - Using the acquired token, a client is able to access REST APIs that are protected with the JWTProvider federation provider.
- TLS Client Certificate Provider
  - enables establishing the user based on the client provided TLS certificate.
  - The user will be the DN from the certificate.

* not all components are supported

HORTONWORKS

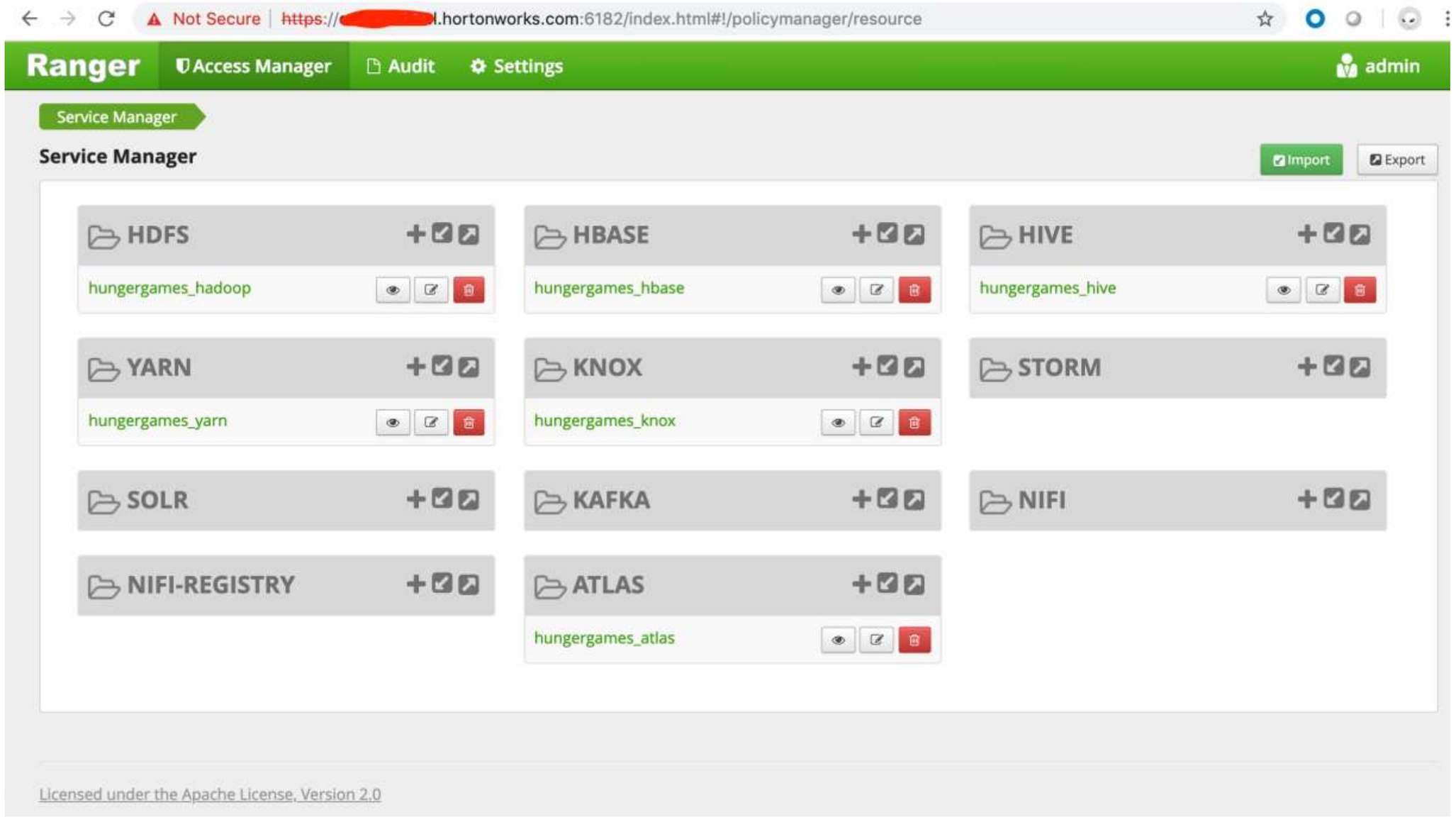
Authorization

Authorization Support in Knox

- Service Level Authorization
- Service based ACL definition – simple Knox authorization
- Apache Ranger Integration
  - for enterprise/centralized authorization policies
  - for finer grained policies on services across the platform (service level for Knox)

HORTONWORKS

Apache Ranger Centralized Policy Management



HORTONWORKS

Audit

Audit – Centralized Logging

- All Service access through the Knox Gateway is Audited with detailed Entries for:
  - Original Access Request
  - Authentication Event
  - Identity Assertion: principal mapping, group lookup
  - Authorization Event
  - Dispatch to backend service

Apache Knox to the Rescue…

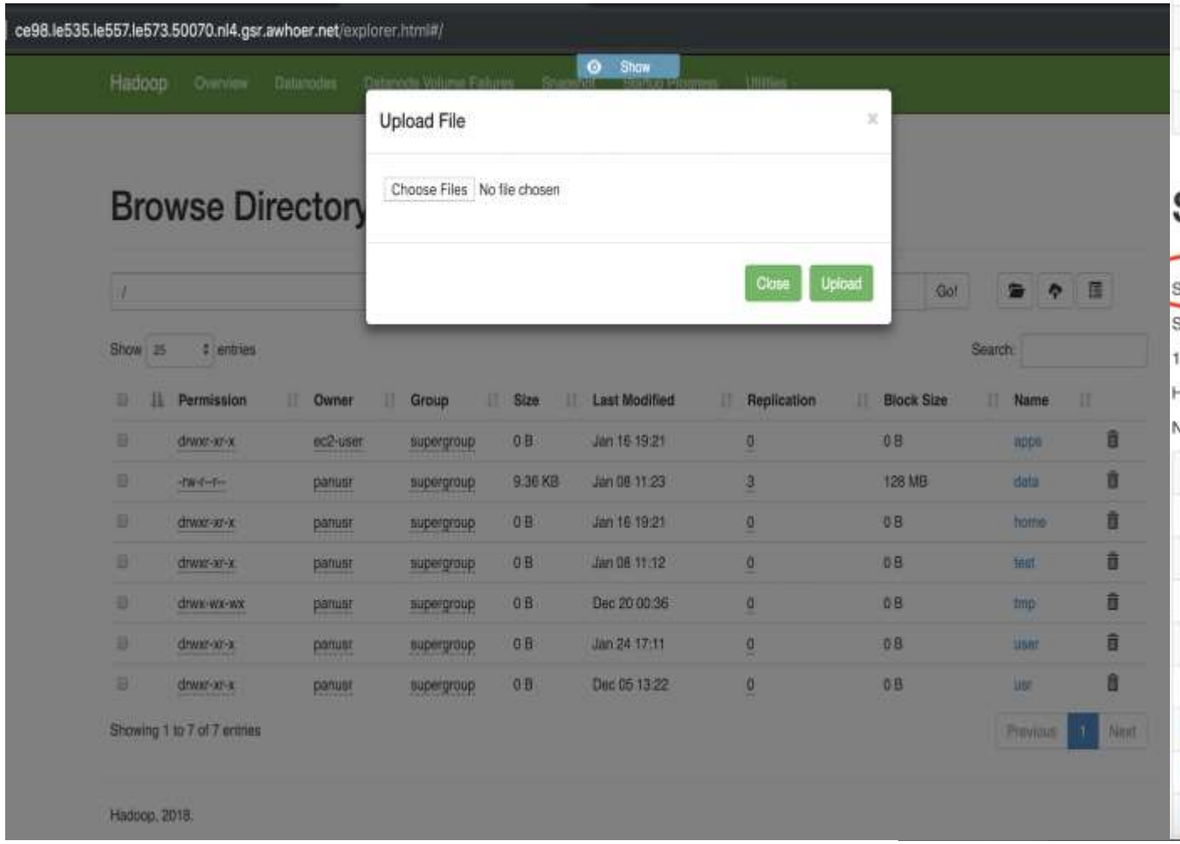
HORTONWORKS®

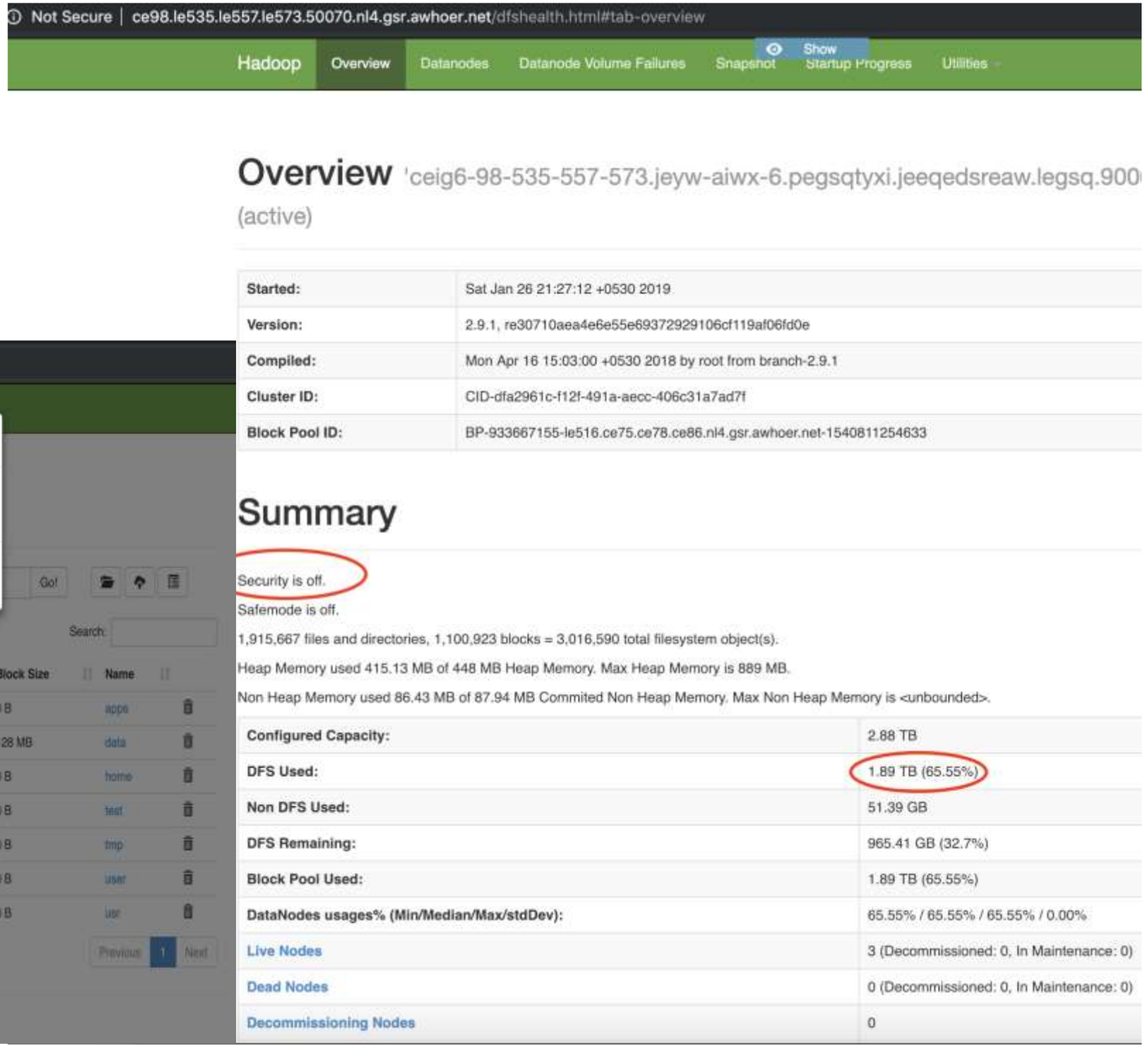
Knox defense against Web Attack Vectors

- Which category Knox has protection against?
  - Security headers
    - XFS
    - CSRF (limited to REST APIs)
    - SSL Downgrade
    - MIME Sniffing
    - XSS
  - Open Redirect – validated redirect & forwards
    - Whitelisting

HORTONWORKS

Let's Protect with Apache Knox





HORTONWORKS

Limit Attack Surface

- Only expose Knox Port through Cloud VPC
- All other Service Access goes through Knox

Add Authenticated Access

- Without Knox or Kerberos enablement, there is no Authentication.
- Proxy Access to the UI through Knox Gateway require HTTP BASIC Auth against LDAP/AD
- Calls to WebHDFS URLs will include the Authenticated User as a user.name Query Param
- Only Authorized Users will have access.

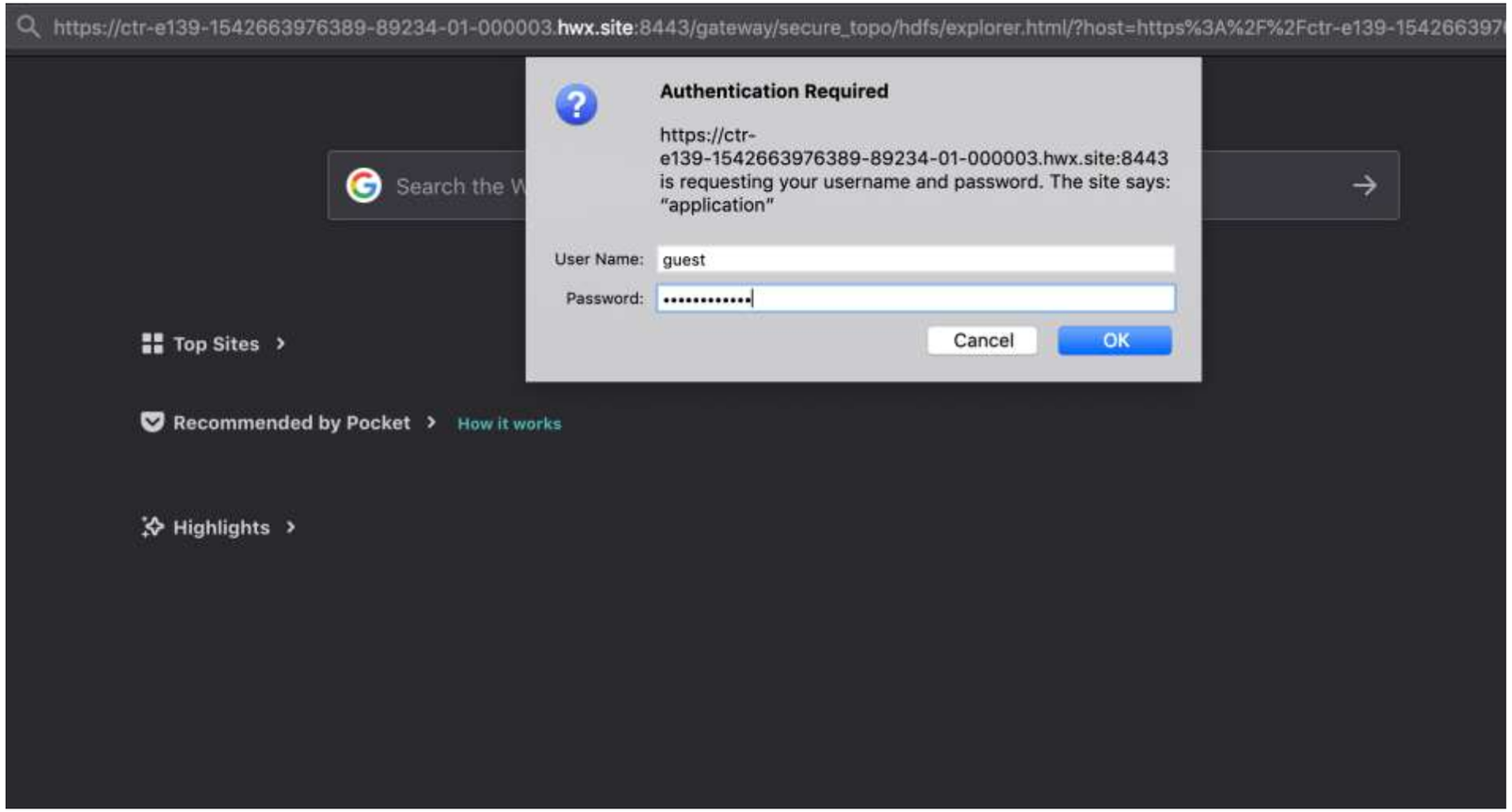
HORTONWORKS

Add Service Level Authorization

- Enable Authorization Providers e.g. Ranger
- Restrict Access to WebHDFS to Users in certain LDAP groups
- Only a subset of your Authenticated Users will have Access to WebHDFS

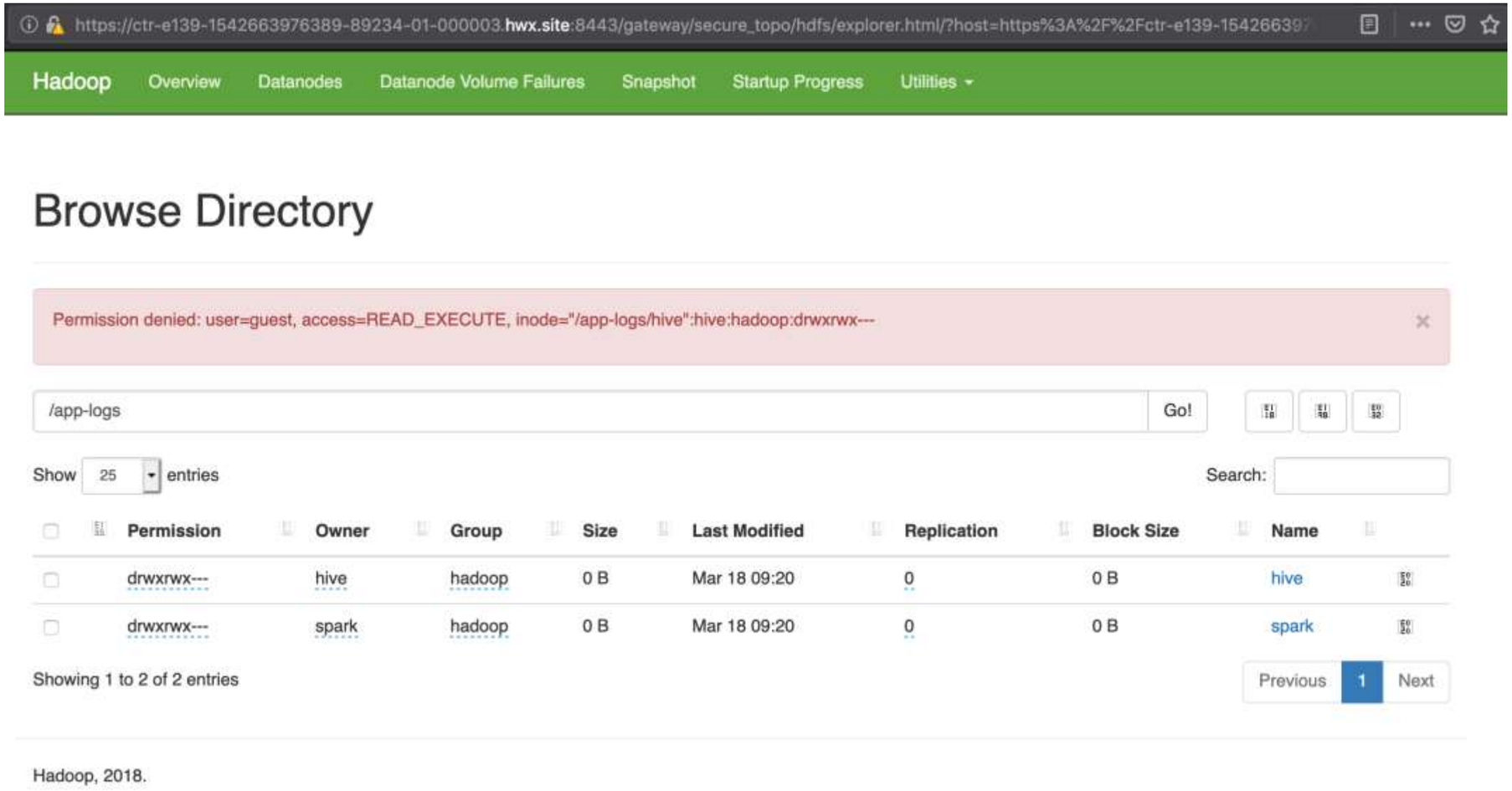
HORTONWORKS

End Result - Authentication



HORTONWORKS

End Result - Authorization



HORTONWORKS

Security Controls - Apache Knox

	Issue Categories (OWASP Top 10 – 2017)	Apache Knox
A1	Injection	No
A2	Broken Authentication	Yes
A3	Sensitive Data Exposure	Yes (Encrypts params & locations)
A4	XML External Entities (XXE)	No
A5	Broken Access Control	Partially (Service level ACLs)
A6	Security Misconfiguration	Partially (Less exposed Services)
A7	XSS	Partially*
A8	Insecure Deserialization	No
A9	Using Components with Known Vulnerabilities	Partially
A10	Insufficient Logging and Monitoring	Partially**

HORTONWORKS

What else you can do?

- Security around your Perimeter
  - Web Application Firewall
  - Intrusion Detection & Prevention System
  - Network (Packet filtering) Firewall
  - API Gateway
- Hardening everything you can
  - O/S - Secure PAM Configuration, SELinux, remove unused packages, disable unused services, etc.
  - Application Server – Jetty, Tomcat, etc.
  - Database Server – enable SSL, restrict IP addresses, etc.
  - KDC Server – master_key_type, supported_enctypes, ticket_lifetime, etc.
  - Set ***.proxyuser.*.hosts/groups** only to needed hosts/groups

Take Away

- Kerberos Encapsulation – Web Browser friendly.
- Knox - Eliminates need for developing central Authentication solution for Hadoop ecosystem
- Provides ready integration with Enterprise Authentication mechanisms like AD, Okta, etc.
- Prevents against several OWASP Top 10 threats

HORTONWORKS


Questions?

Thank you