MANAGING ENTERPRISE USERS IN HADOOP ECOSYSTEM

Sailaja Polavarapu
Staff Software Engineer
Hortonworks
spolavarapu@apache.org

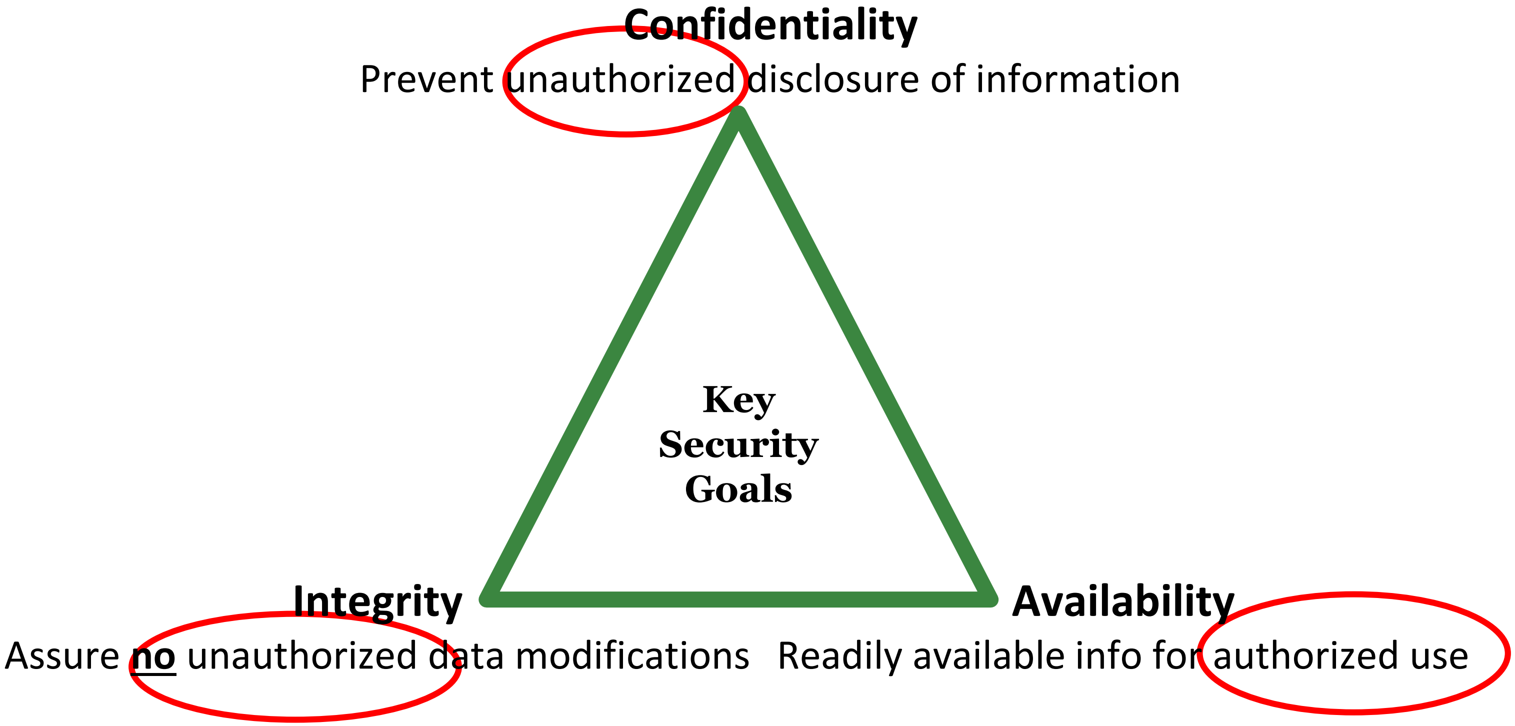
Velmurugan Periasamy
Director, Engineering
Hortonworks
vel@apache.org

Dataworks Summit 2018 San Jose

Agenda

- Introduction
- Enterprise User Management in Hadoop environment
- Enabling technologies
- Integrating Hadoop cluster with Multiple LDAP stores
- Key takeaways
- Demo

Managing Enterprise Users - Why?



Confidentiality

Prevent unauthorized disclosure of information

Key Security Goals

Integrity

Assure **no** unauthorized data modifications

Availability

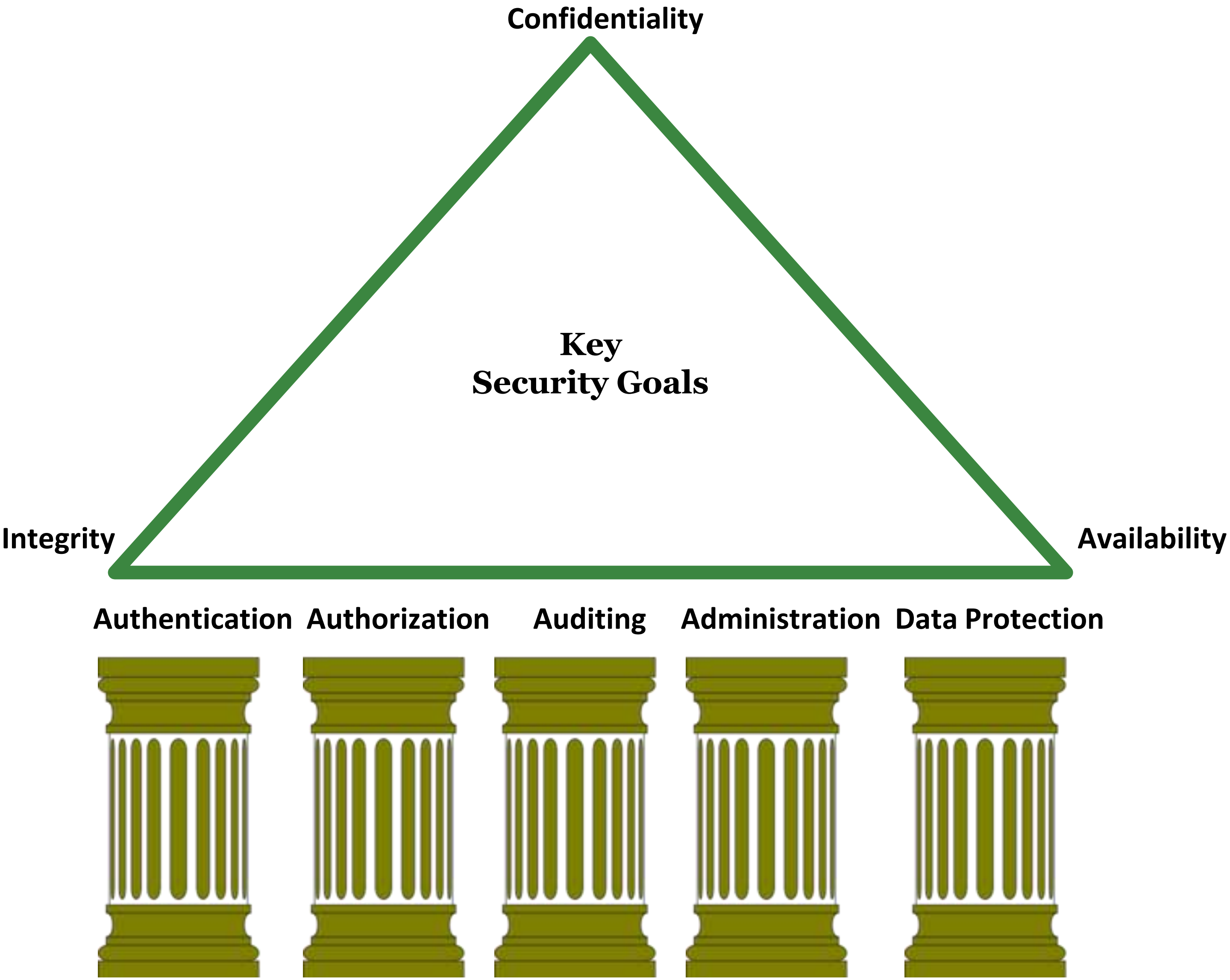
Readily available info for authorized use

Managing Enterprise Users - Why?

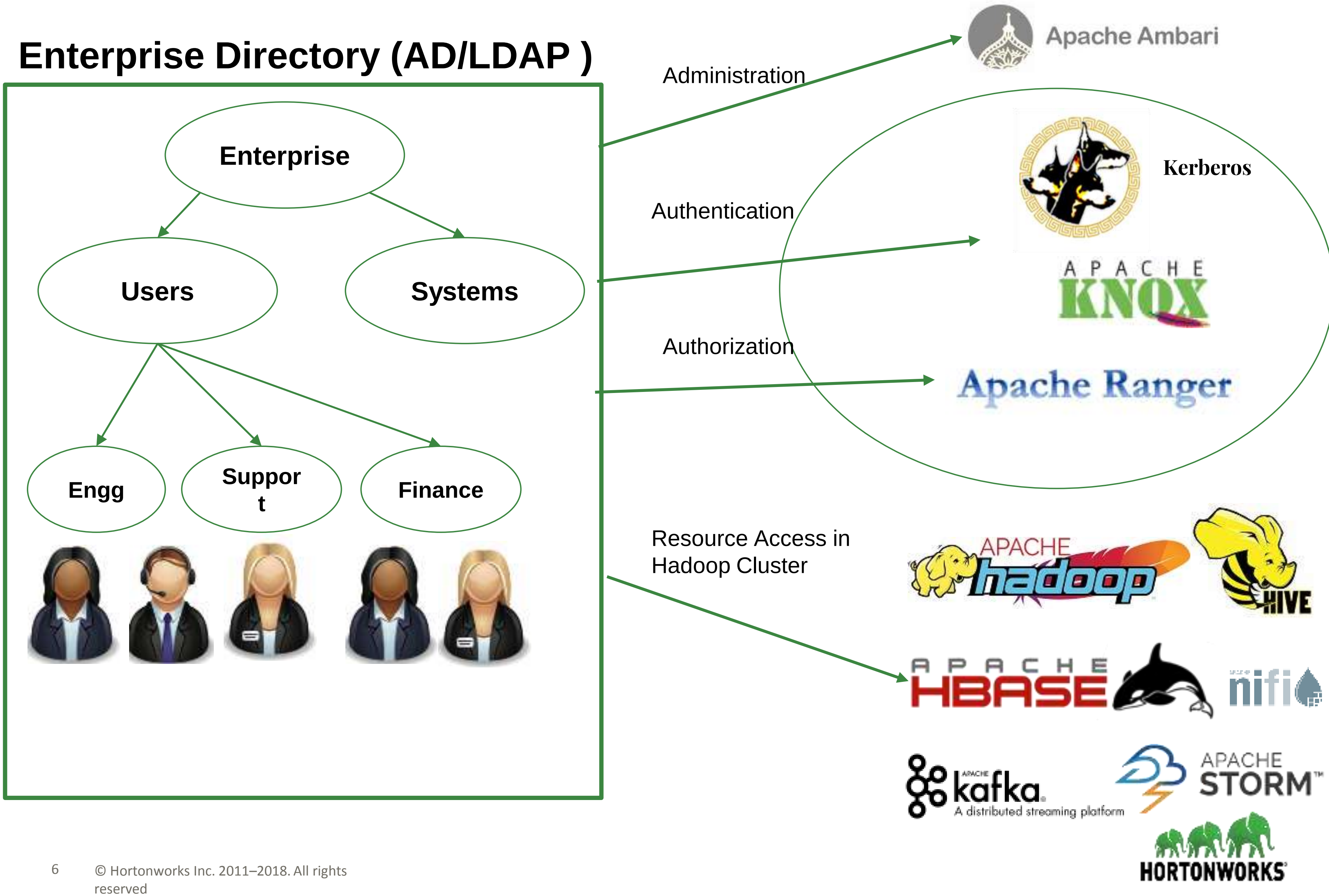
- Cost of Average Data breach - **$3.62M**
- Factor in the loss in business continuity, regulatory/PR/legal issues, damage to the brand etc.
- Security is serious business
- Enterprise Users Management is critical

Source: https://www.ibm.com/security/data-breach

Managing Enterprise Users - Foundations



Managing Enterprise Users in Hadoop Env



Enterprise Directory (AD/LDAP)

Enterprise

Users

Systems

Engg

Suppor
t

Finance

Administration

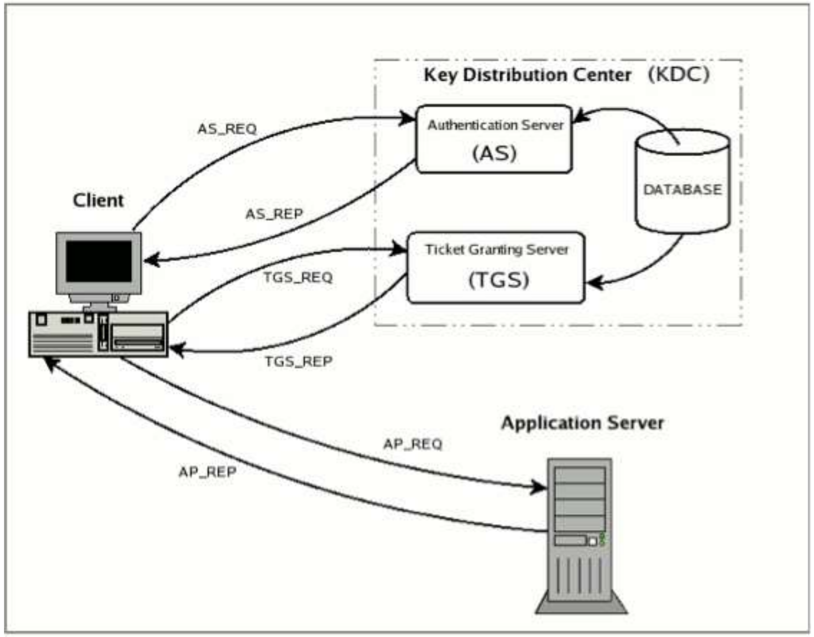
Authentication

Authorization

Resource Access in
Hadoop Cluster

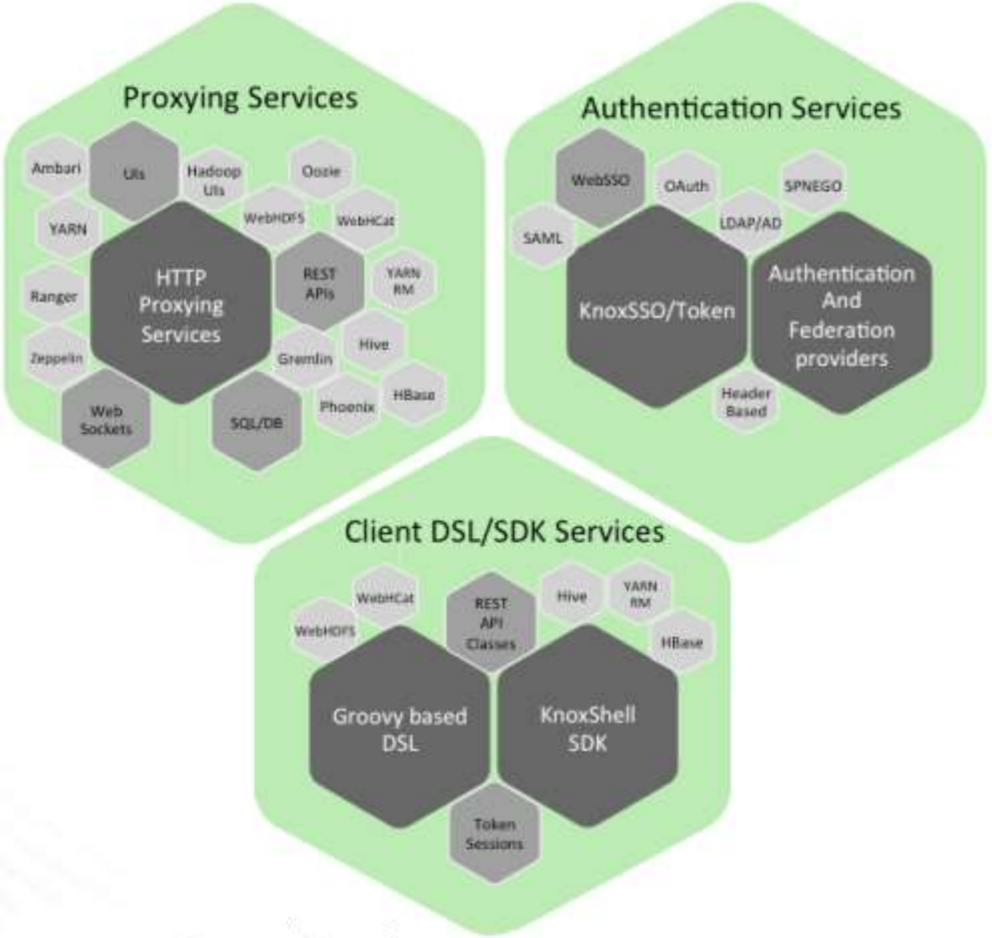
Kerberos

- Open Standard Authentication Protocol
- Developed by MIT for distributed systems
- Kerberos is MUST for Hadoop Env
  - Authenticating to KDC provides access for all services in REALM
- Directory Server Integration



Apache Knox

- Centralized authentication with SSO
- Complements, does not replace Kerberos
- Single access point for all REST and HTTP interactions with Apache Hadoop clusters
- Directory server integration



Proxying Services

Ambari
UIs
Hadoop UIs
Oozie
YARN
WebHDFS
WebHCat
HTTP Proxying Services
REST APIs
YARN RM
Ranger
Hive
Zeppelin
Gremlin
HBase
Web Sockets
SQL/DB
Phoenix

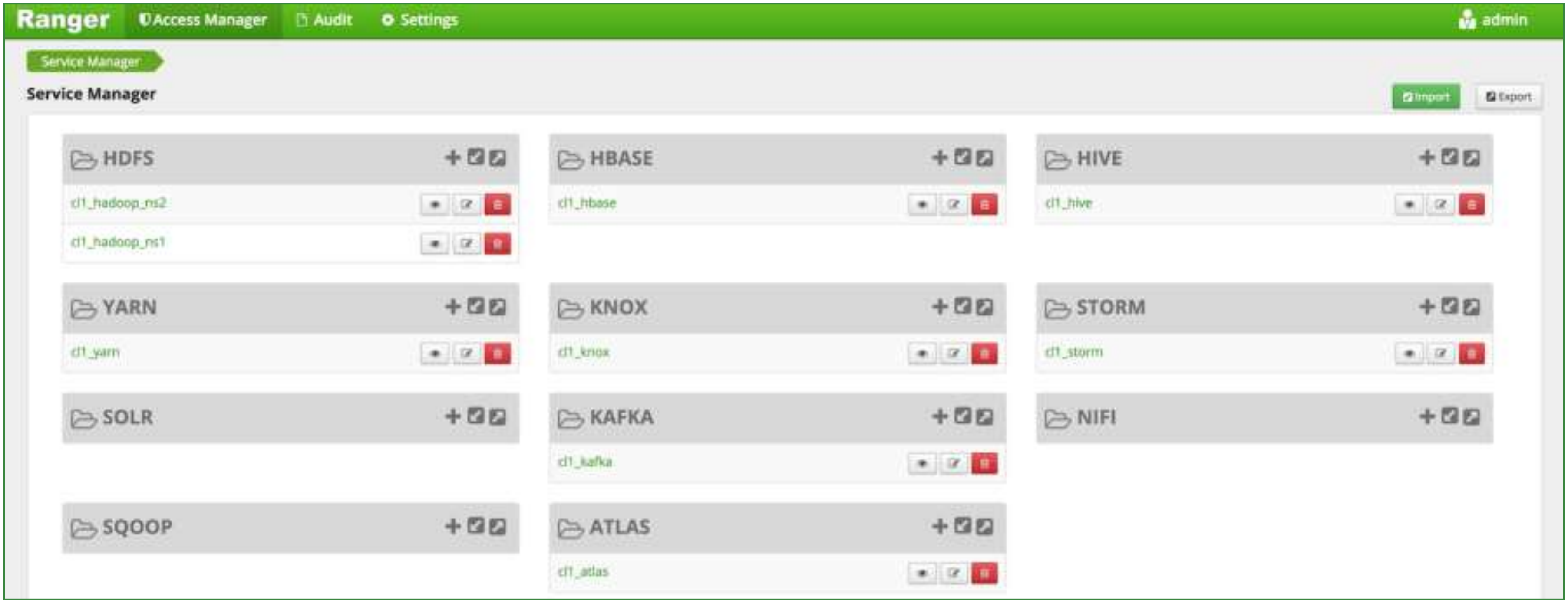
Authentication Services

WebSSO
OAuth
SPNEGO
SAML
LDAP/AD
KnoxSSO/Token
Authentication And Federation providers
Header Based

Client DSL/SDK Services

WebHCat
REST API Classes
Hive
YARN RM
WebHDFS
HBase
Groovy based DSL
KnoxShell SDK
Token Sessions

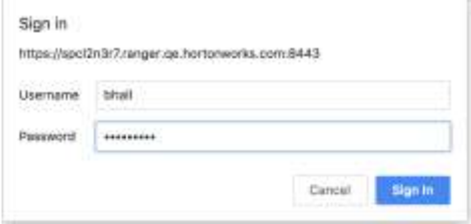
Apache Ranger



- Centralized authorization & auditing for hadoop ecosystem
- Fine grained access control with flexible ABAC models
- Centralized security policy administration
- Directory server integration

User Management in Hadoop: Drill Down

Authentication



Sign in

https://spcl2n3r7.ranger.qe.hortonworks.com:8443

Username: bhall

Password: *********

Authorization

User accessing hadoop resources

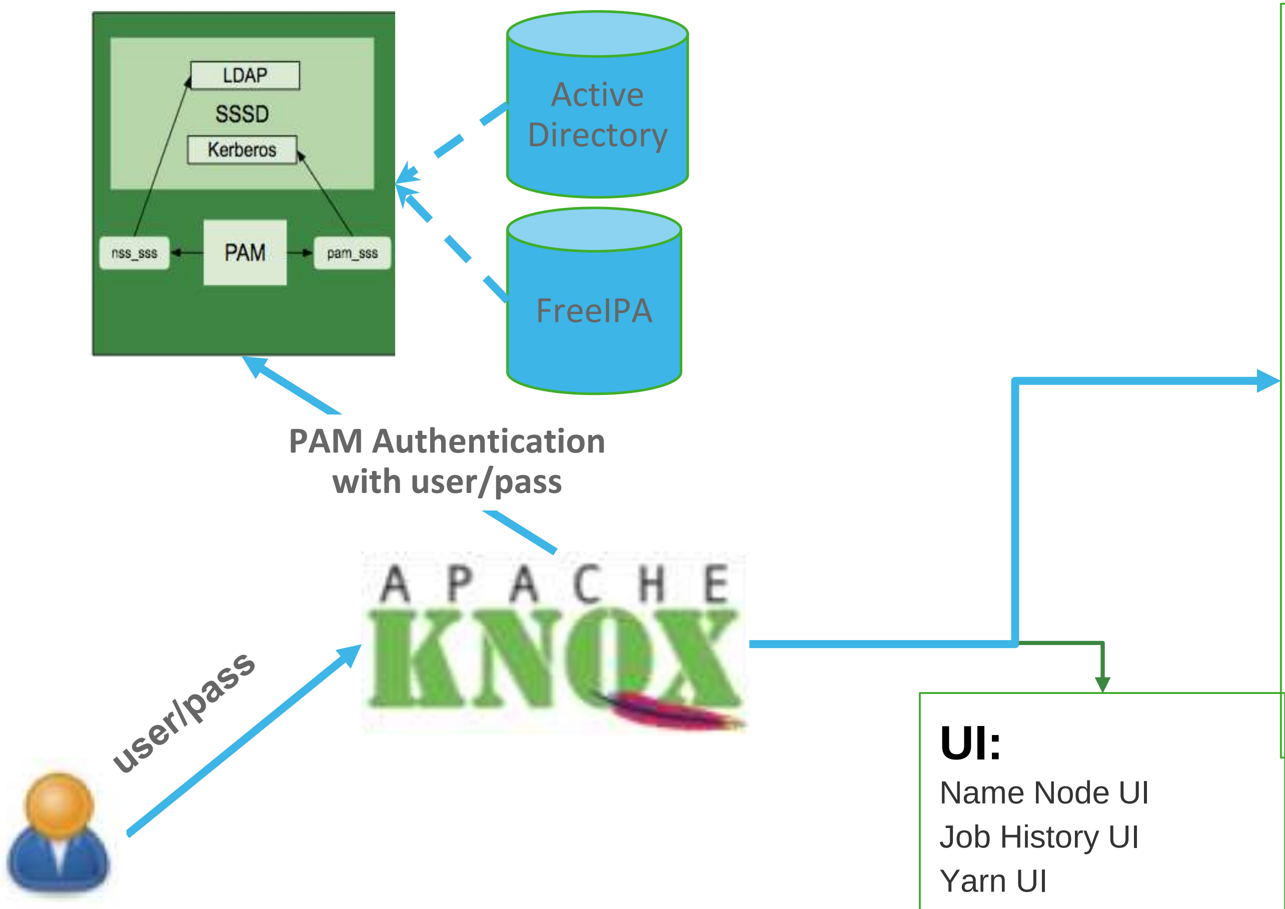
Requirements for User Management

- Understanding your deployment
  - What kind of directory server(s): Active Directory, OpenLdap server, etc…?
  - OS version of the hadoop cluster nodes: CentOS, Ubuntu, etc...
  - User group mapping on hadoop clusters: using SSSD, core-site.xml, manual, etc…
  - Authentication mechanism: Kerberos, Knox gateway, etc...
  - Authorization policies to be configured at user level or group level?

User Management in Knox

- Supports Authentication using either LDAP or Federation provider
  – Active Directory/LDAP
  – SPNEGO/Kerberos
  – PAM (Pluggable Authentication Module)

Authentication flow using Knox



Active Directory

FreeIPA

PAM Authentication with user/pass

user/pass

https://localhost:8443/gateway/default/webhdfs/v1/?op=LISTSTATUS

Services:

Ambari
WebHDFS (HDFS)
Yarn RM
Stargate (Apache HBase)
Apache Oozie
Apache Hive/JDBC
Apache Hive WebHCat (Templeton)
Apache Storm
Apache Tinkerpop - Gremlin
Apache Avatica/Phoenix
Apache SOLR
Apache Livy (Spark REST Service)
Kafka REST Proxy

UI:

Name Node UI
Job History UI
Yarn UI
Apache Oozie UI
Apache HBase UI
Apache Spark UI
Apache Ambari UI
Apache Ranger Admin Console
Apache Zeppelin

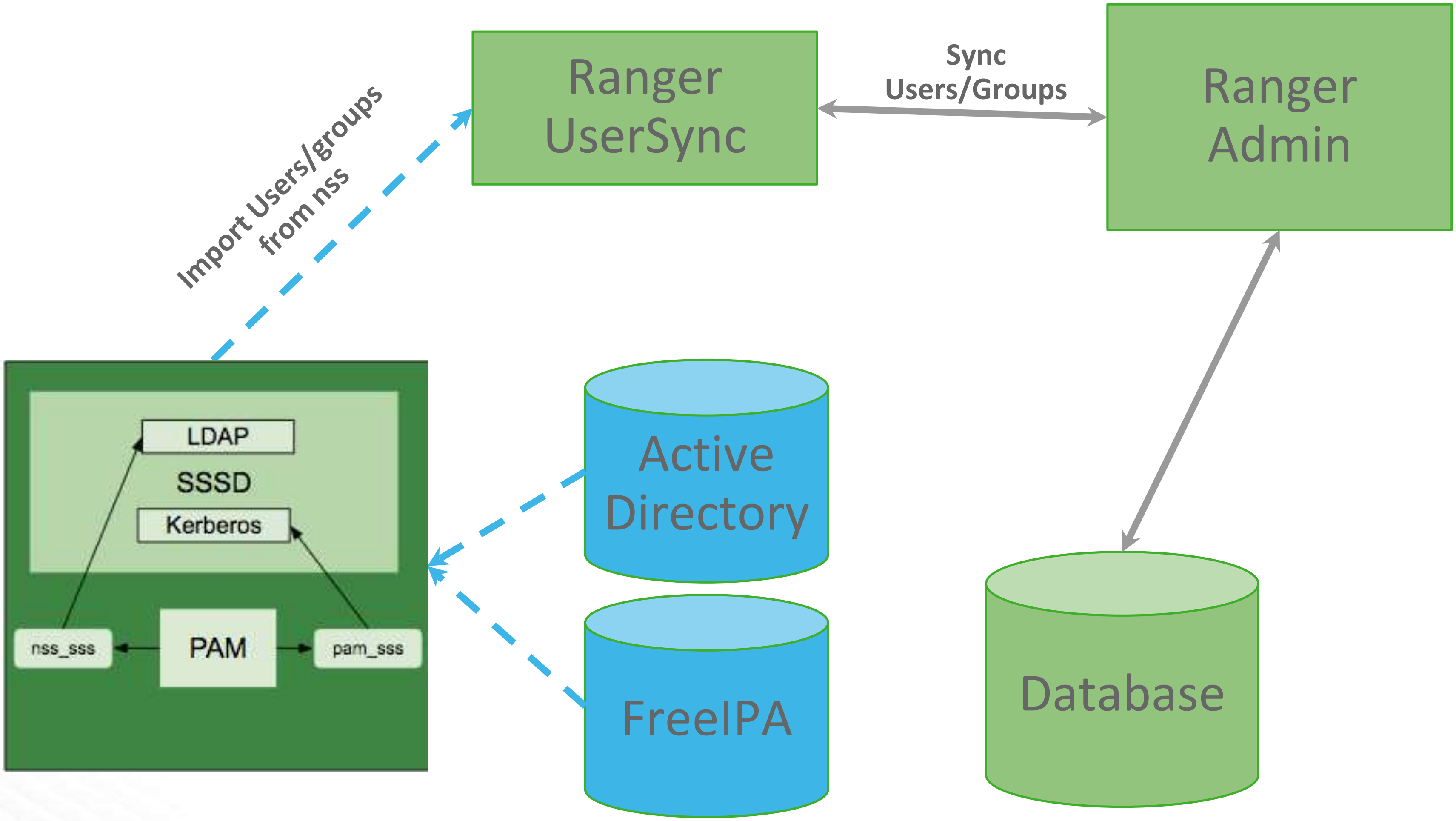
User Management in Ranger

- Users and Groups in Ranger are used for
  - Configuring policies
  - Authorization or access control
  - Auditing
- Users and Groups management in ranger is handled by UserSync module
- UserSync modules key responsiblities
  - Have Ranger access to enterprise users and groups during policy creation
  - Interact with directory servers and provide users, groups, and membership updates to ranger
  - Provides flexible configuration options
  - Provides Usersync Audits (Introduced in HDP 3.0)

User sources

- AD/LDAP
  - Syncs users and groups from LDAP Organizational Units (OU)
- Unix Native Users
  - Syncs users and groups from /etc/passwd and /etc/group files
  - Sync users and groups provided by NSS (Name Service Switch) (Introduced in HDP 3.0)
- File Sources
  - Syncs users and groups from a file specified in the configuration.
  - Supports many file formats like - CSV, JSON, etc...

User/Group Synchronization in Ranger



Ranger UserSync

Sync Users/Groups

Ranger Admin

Import Users/groups from nss

Active Directory

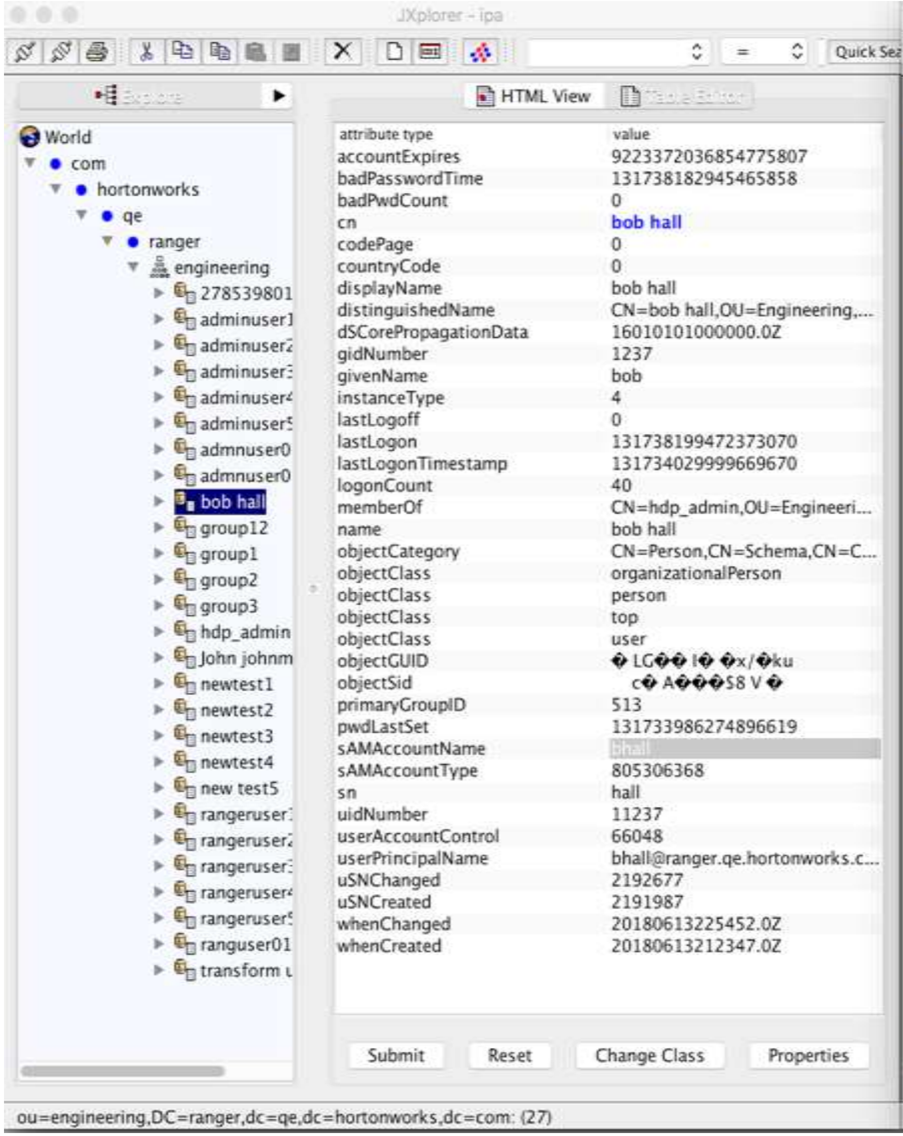
FreeIPA

Database

Use case

- LDAP requirements
  - Multiple LDAP (IBM Tivoli, Lotus Notes Dominoes, Active Directory On-prem, Azure Active Directory)
  - Single Enterprise Data Lake Cluster across multiple countries

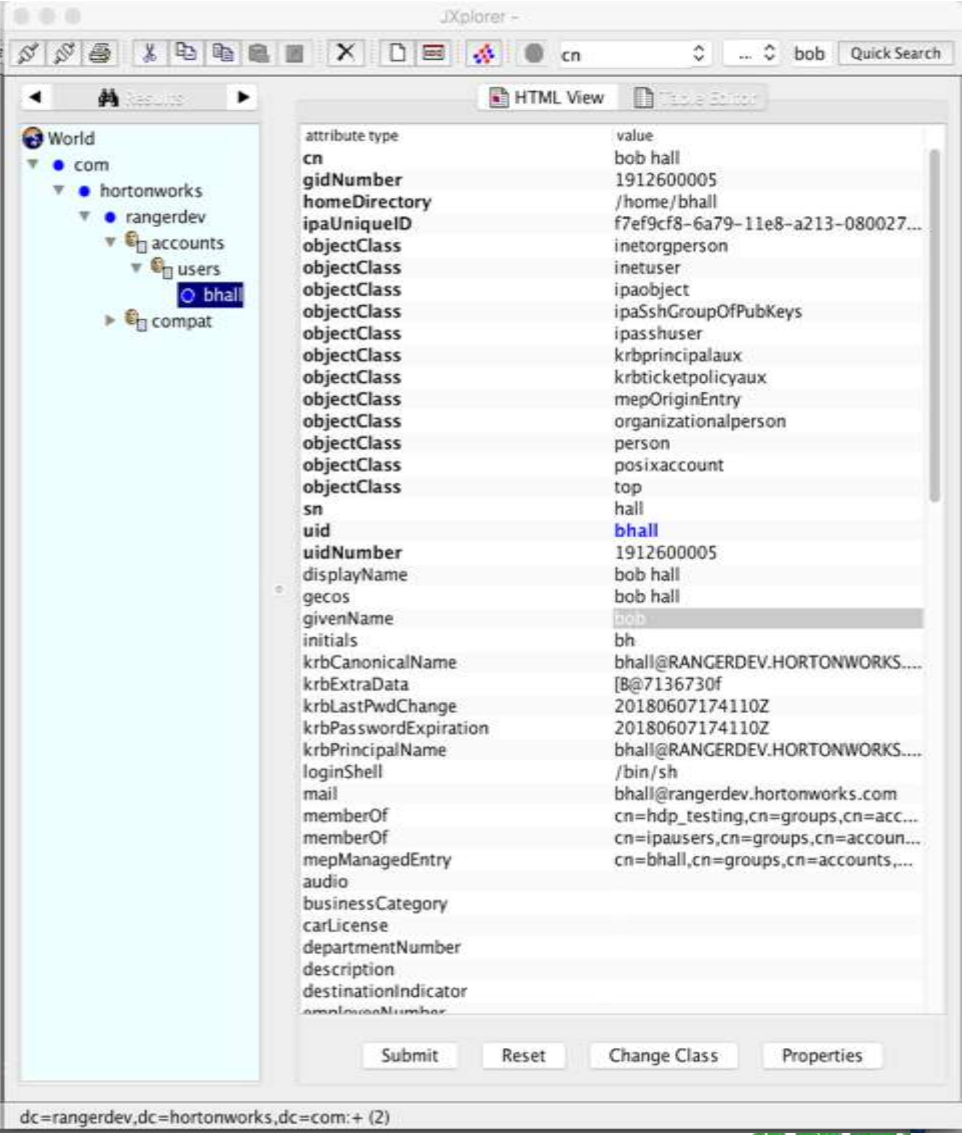
Sample User from two different domains



- World
  - com
    - hortonworks
      - qe
        - ranger
          - engineering
            - 278539801
            - adminuser1
            - adminuser2
            - adminuser3
            - adminuser4
            - adminuser5
            - admnuser0
            - admnuser0
            - bob hall
            - group12
            - group1
            - group2
            - group3
            - hdp_admin
            - John johnm
            - newtest1
            - newtest2
            - newtest3
            - newtest4
            - new test5
            - rangeruser1
            - rangeruser2
            - rangeruser3
            - rangeruser4
            - rangeruser5
            - ranguser01
            - transform u

attribute type	value
accountExpires	9223372036854775807
badPasswordTime	131738182945465858
badPwdCount	0
cn	bob hall
codePage	0
countryCode	0
displayName	bob hall
distinguishedName	CN=bob hall,OU=Engineering,...
dSCorePropagationData	16010101000000.0Z
gidNumber	1237
givenName	bob
instanceType	4
lastLogoff	0
lastLogon	131738199472373070
lastLogonTimestamp	131734029999669670
logonCount	40
memberOf	CN=hdp_admin,OU=Engineeri...
name	bob hall
objectCategory	CN=Person,CN=Schema,CN=C...
objectClass	organizationalPerson
objectClass	person
objectClass	top
objectClass	user
objectGUID	� LG�� I� �x/�ku
objectSid	c� A���$8 V �
primaryGroupID	513
pwdLastSet	131733986274896619
sAMAccountName	bhall
sAMAccountType	805306368
sn	hall
uidNumber	11237
userAccountControl	66048
userPrincipalName	bhall@ranger.qe.hortonworks.c...
uSNChanged	2192677
uSNCreated	2191987
whenChanged	20180613225452.0Z
whenCreated	20180613212347.0Z

ou=engineering,DC=ranger,dc=qe,dc=hortonworks,dc=com: (27)

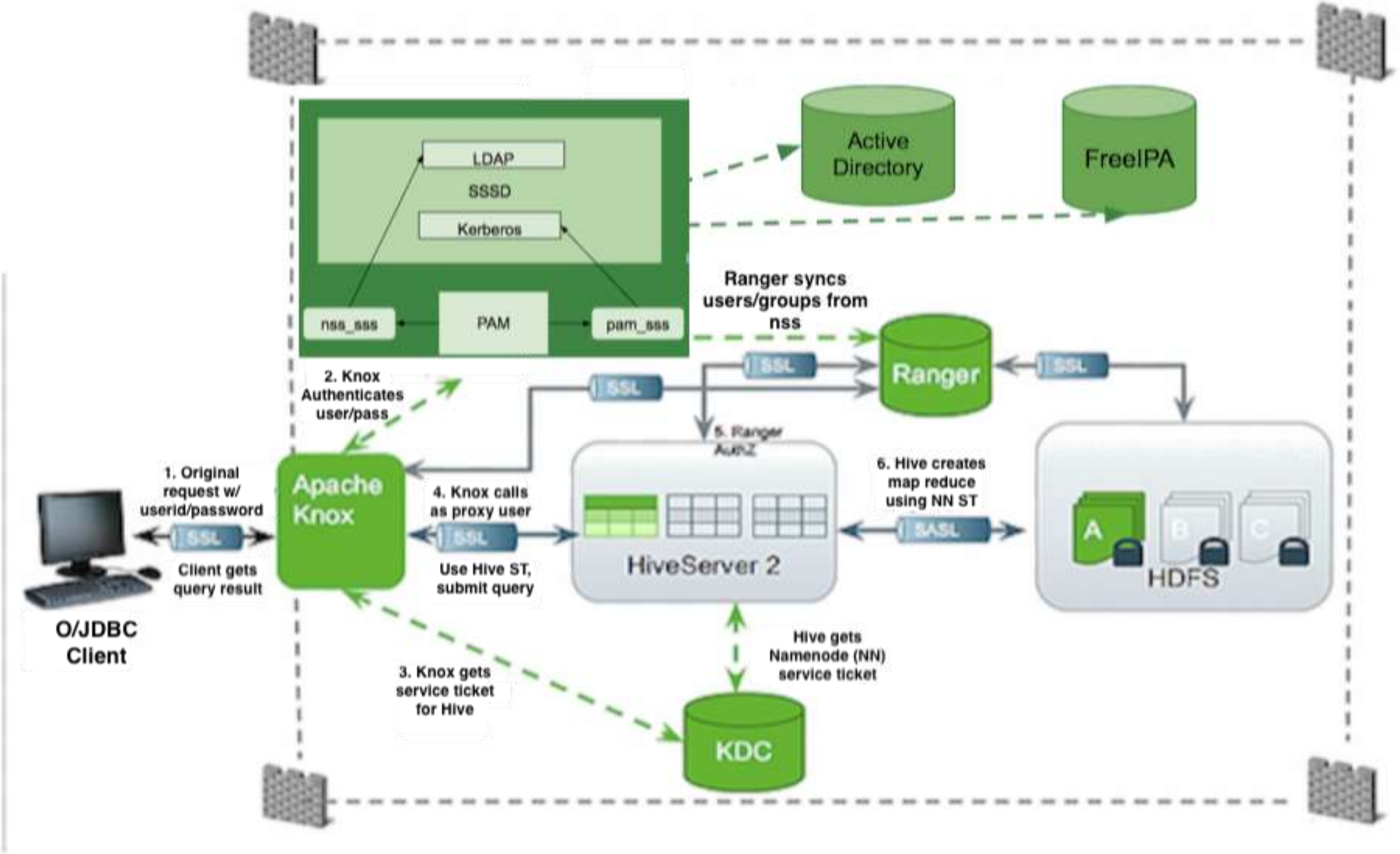


- World
  - com
    - hortonworks
      - rangerdev
        - accounts
          - users
            - bhall
        - compat

attribute type	value
cn	bob hall
gidNumber	1912600005
homeDirectory	/home/bhall
ipaUniqueID	f7ef9cf8-6a79-11e8-a213-080027...
objectClass	inetorgperson
objectClass	inetuser
objectClass	ipaobject
objectClass	ipaSshGroupOfPubKeys
objectClass	ipasshuser
objectClass	krbprincipalaux
objectClass	krbticketpolicyaux
objectClass	mepOriginEntry
objectClass	organizationalperson
objectClass	person
objectClass	posixaccount
objectClass	top
sn	hall
uid	bhall
uidNumber	1912600005
displayName	bob hall
gecos	bob hall
givenName	bob
initials	bh
krbCanonicalName	bhall@RANGERDEV.HORTONWORKS....
krbExtraData	[B@7136730f
krbLastPwdChange	20180607174110Z
krbPasswordExpiration	20180607174110Z
krbPrincipalName	bhall@RANGERDEV.HORTONWORKS....
loginShell	/bin/sh
mail	bhall@rangerdev.hortonworks.com
memberOf	cn=hdp_testing,cn=groups,cn=acc...
memberOf	cn=ipausers,cn=groups,cn=accoun...
mepManagedEntry	cn=bhall,cn=groups,cn=accounts,...
audio	
businessCategory	
carLicense	
departmentNumber	
description	
destinationIndicator	
employeeNumber	

dc=rangerdev,dc=hortonworks,dc=com:+ (2)

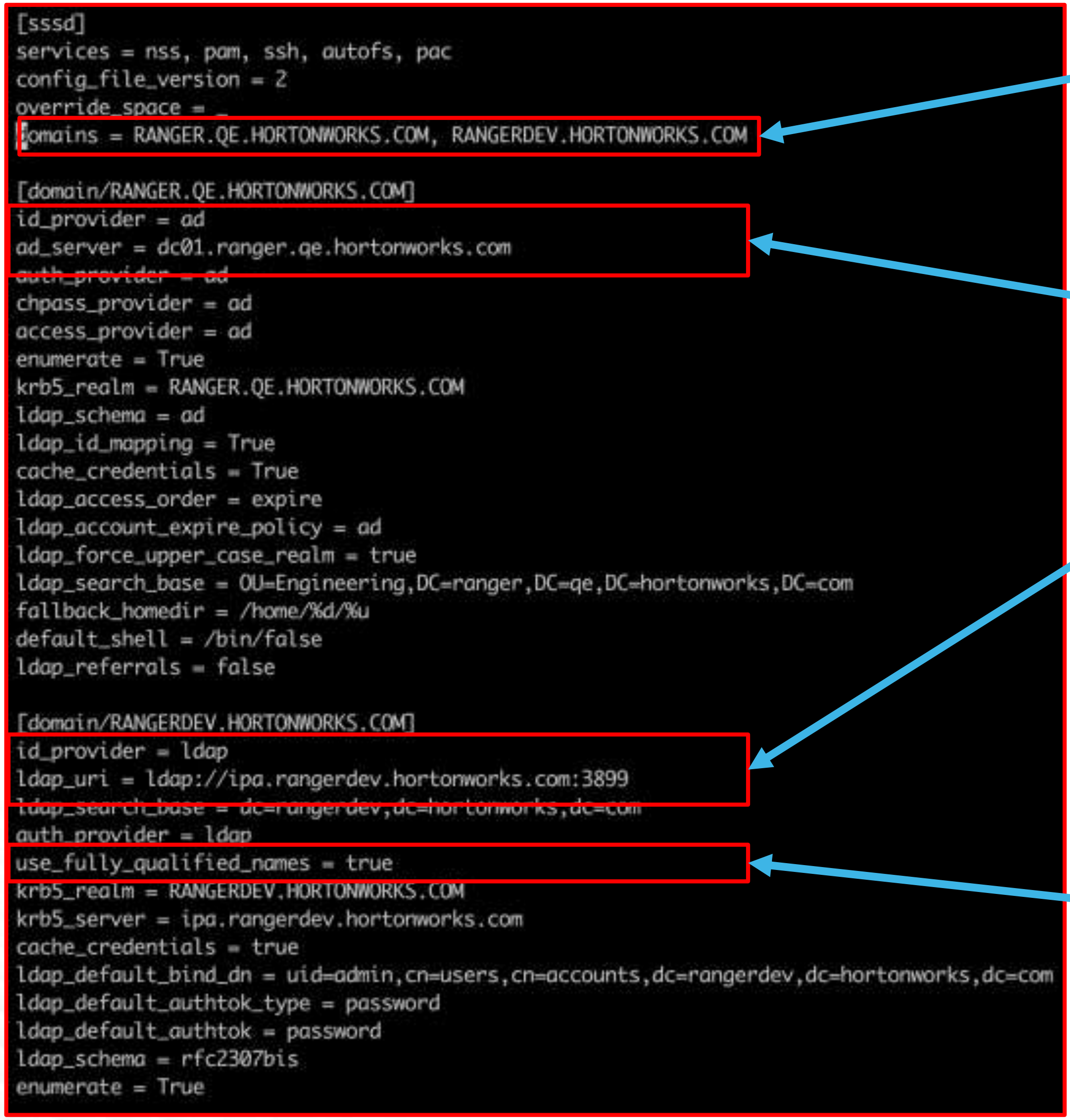
Reference Architecture


LDAP
SSSD
Kerberos
nss_sss
PAM
pam_sss
Active Directory
FreeIPA
Ranger syncs users/groups from nss
Ranger
SSL
SSL
SSL
2. Knox Authenticates user/pass
5. Ranger AuthZ
1. Original request w/ userid/password
SSL
Client gets query result
Apache Knox
4. Knox calls as proxy user
SSL
Use Hive ST, submit query
HiveServer 2
6. Hive creates map reduce using NN ST
SASL
A
B
C
HDFS
O/JDBC Client
3. Knox gets service ticket for Hive
Hive gets Namenode (NN) service ticket
KDC

Steps Involved

- Setting up secure cluster (HDP version used for this demo is 3.0 and centos7)
- Active directory and FreeIPA setup
- Setting up SSSD with PAM and NSS services
- Configuring Knox with PAM+SSSD for authentication
- Configuring Ranger admin & usersync with PAM + SSSD
- Configuring services like hdfs & hive with kerberos authentication

SSSD configuration with AD and FreeIPA servers



```
[sssd]
services = nss, pam, ssh, autofs, pac
config_file_version = 2
override_space = _
domains = RANGER.QE.HORTONWORKS.COM, RANGERDEV.HORTONWORKS.COM

[domain/RANGER.QE.HORTONWORKS.COM]
id_provider = ad
ad_server = dc01.ranger.qe.hortonworks.com
auth_provider = ad
chpass_provider = ad
access_provider = ad
enumerate = True
krb5_realm = RANGER.QE.HORTONWORKS.COM
ldap_schema = ad
ldap_id_mapping = True
cache_credentials = True
ldap_access_order = expire
ldap_account_expire_policy = ad
ldap_force_upper_case_realm = true
ldap_search_base = OU=Engineering,DC=ranger,DC=qe,DC=hortonworks,DC=com
fallback_homedir = /home/%d/%u
default_shell = /bin/false
ldap_referrals = false

[domain/RANGERDEV.HORTONWORKS.COM]
id_provider = ldap
ldap_uri = ldap://ipa.rangerdev.hortonworks.com:3899
ldap_search_base = dc=rangerdev,dc=hortonworks,dc=com
auth_provider = ldap
use_fully_qualified_names = true
krb5_realm = RANGERDEV.HORTONWORKS.COM
krb5_server = ipa.rangerdev.hortonworks.com
cache_credentials = true
ldap_default_bind_dn = uid=admin,cn=users,cn=accounts,dc=rangerdev,dc=hortonworks,dc=com
ldap_default_authtok_type = password
ldap_default_authtok = password
ldap_schema = rfc2307bis
enumerate = True
```

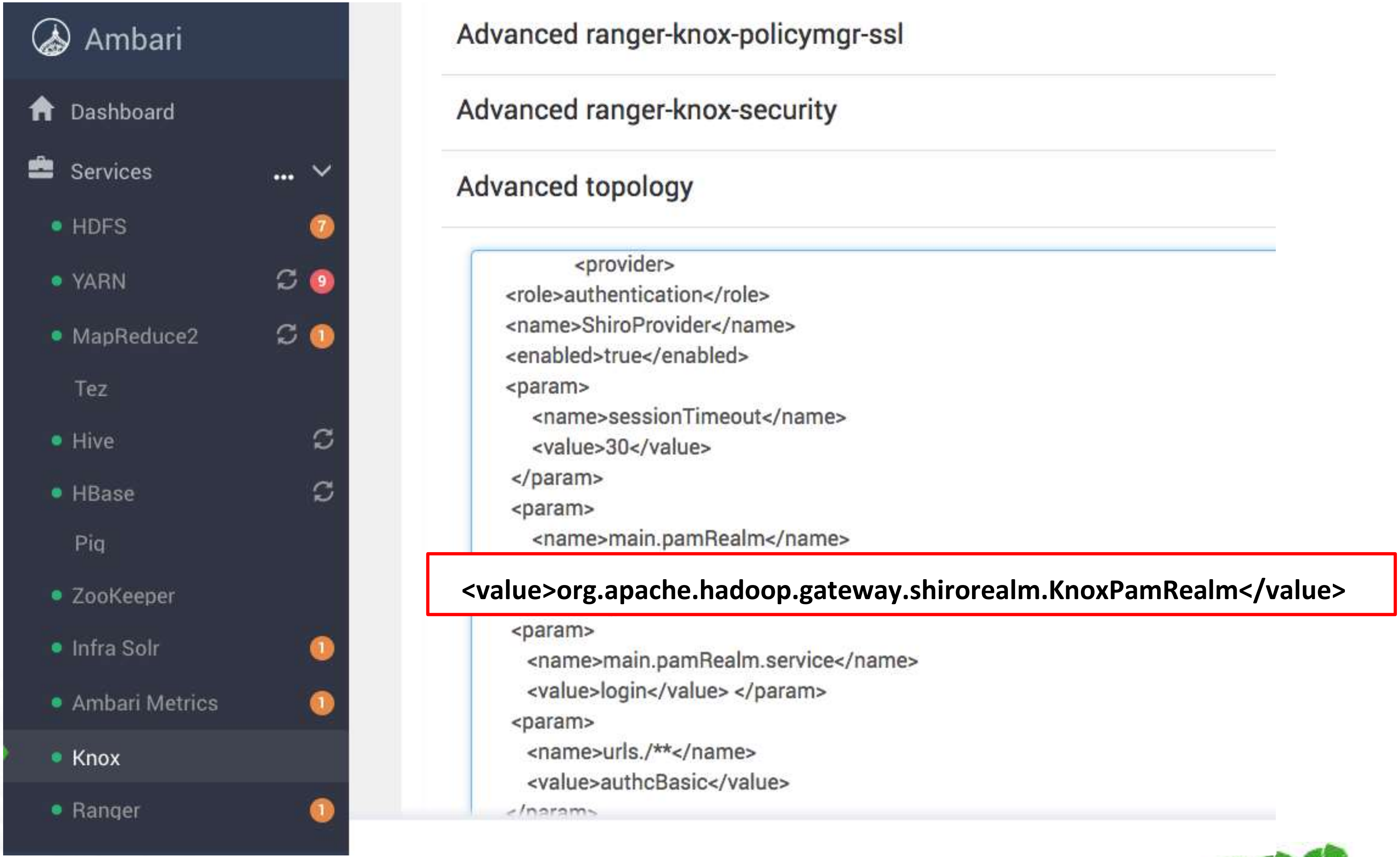
Multiple Domains:
RANGER.QE.HORTONWORKS.COM
RANGERDEV.HORTONWORKS.COM

RANGER.QE.HORTONWORKS.COM is an Active Directory

RANGERDEV.HORTONWORKS.COM is a FreeIPA server

Using Fully qualified names for users and groups from RANGERDEV.HORTONWORKS.COM in order to resolve name conflicts between two domains

Knox topology config for PAM Authentication



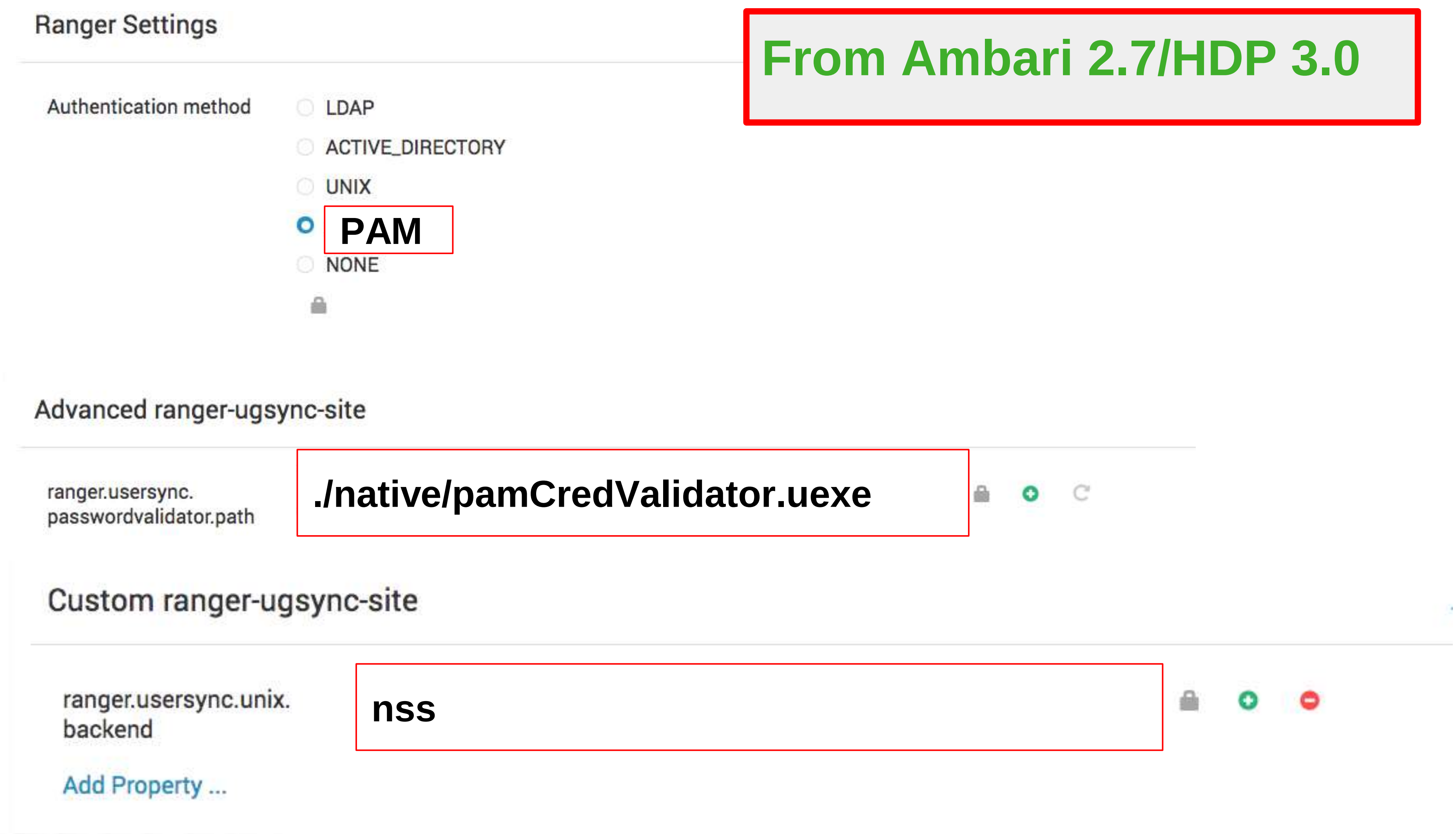
Advanced ranger-knox-policymgr-ssl

Advanced ranger-knox-security

Advanced topology

```
            <provider>
    <role>authentication</role>
    <name>ShiroProvider</name>
    <enabled>true</enabled>
    <param>
        <name>sessionTimeout</name>
        <value>30</value>
     </param>
     <param>
        <name>main.pamRealm</name>
 <value>org.apache.hadoop.gateway.shirorealm.KnoxPamRealm</value>
     <param>
       <name>main.pamRealm.service</name>
       <value>login</value> </param>
     <param>
       <name>urls./**</name>
       <value>authcBasic</value>
    </param>
```

Ranger Configuration for PAM and NSS



From Ambari 2.7/HDP 3.0

Ranger Settings

Authentication method

○ LDAP
○ ACTIVE_DIRECTORY
○ UNIX
● PAM
○ NONE

Advanced ranger-ugsync-site

ranger.usersync.passwordvalidator.path: ./native/pamCredValidator.uexe

Custom ranger-ugsync-site

ranger.usersync.unix.backend: nss

Add Property ...

Demo

Key Takeaways

- Kerberos is essential in Hadoop env
- Integration with enterprise directory critical for seamless security
- Centralized Security Administration
- Ranger for Centralized Authorization and Auditing
- Knox for Centralized REST/HTTP Authentication
- Multiple domains can be integrated via SSSD

Questions?

References

- Apache Ranger -
  http://ranger.apache.org/
  https://hortonworks.com/apache/ranger/
  https://cwiki.apache.org/confluence/display/RANGER
- Apache Knox -
  http://knox.apache.org/
  https://hortonworks.com/apache/knox-gateway/
  https://cwiki.apache.org/confluence/display/KNOX/
- Configuring Knox with SSSD + PAM -
  https://cwiki.apache.org/confluence/pages/viewpage.action?pageId=66854729
  https://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.6.0/bk_security/content/setting_up_pam_authentication.html
- Configuring Ranger admin with SSSD + PAM -
  https://issues.apache.org/jira/browse/RANGER-842
- Configuring Ranger Usersync with SSSD + PAM -
  https://issues.apache.org/jira/browse/RANGER-827

Join Apache Ranger Community

user@ranger.apache.org
dev@ranger.apache.org