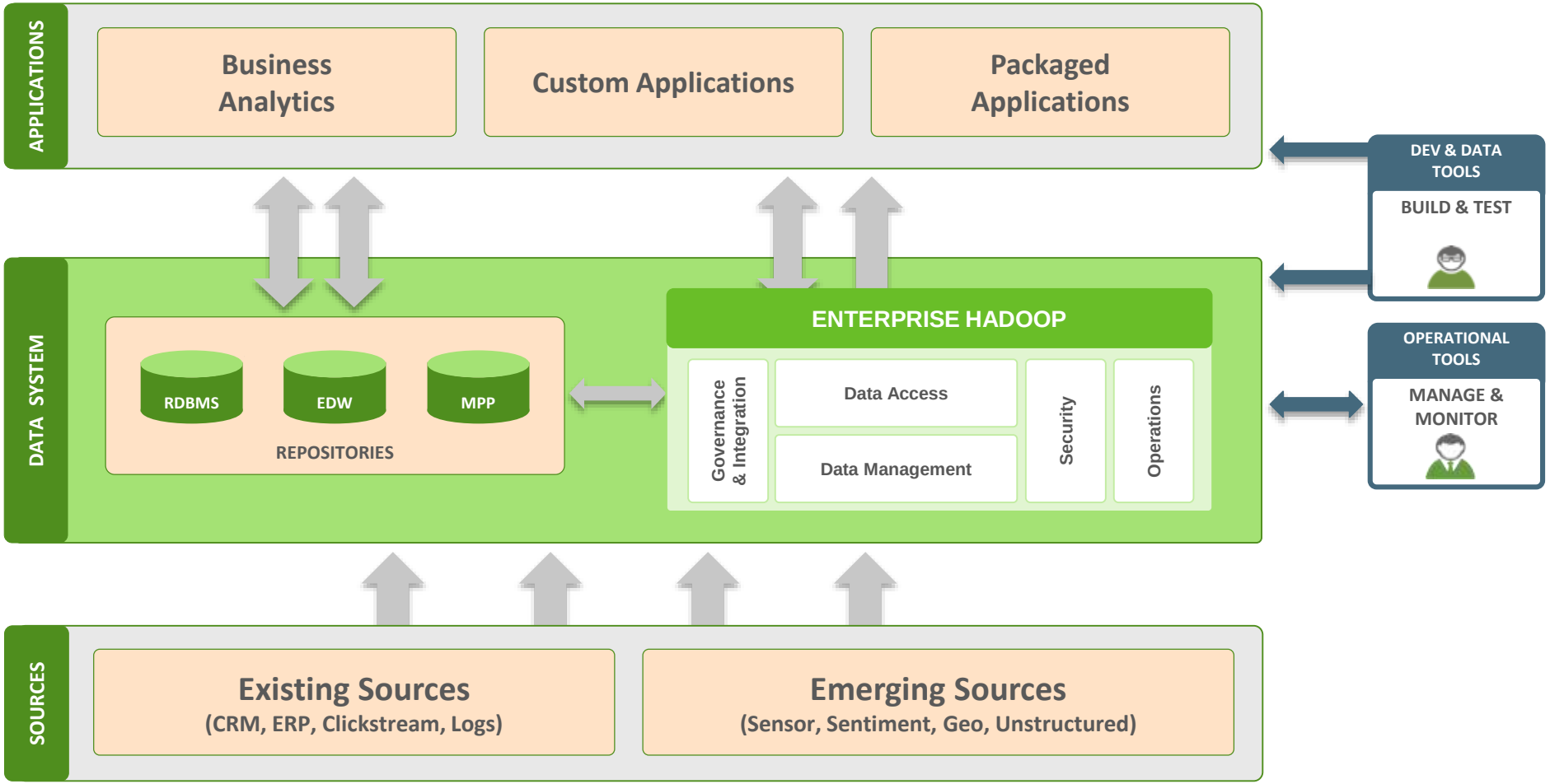
Securing Hadoop Data Lake

Hortonworks. We do Hadoop.

Agenda

- **Security Approach within Hadoop**
- **Security Pillars**
- **Workshops**
- **Questions**

A Modern Data Architecture



Core Capabilities of Enterprise Hadoop

PRESENTATION & APPLICATION

Enable both existing and new application to provide value to the organization

ENTERPRISE MGMT & SECURITY

Empower existing operations and security tools to manage Hadoop

GOVERNANCE & INTEGRATION

Load data and manage according to policy

DATA ACCESS

Access your data simultaneously in multiple ways (batch, interactive, real-time)

Store and process all of your Corporate Data Assets

DATA MANAGEMENT

SECURITY

Provide layered approach to security through Authentication, Authorization, Accounting, and Data Protection

OPERATIONS

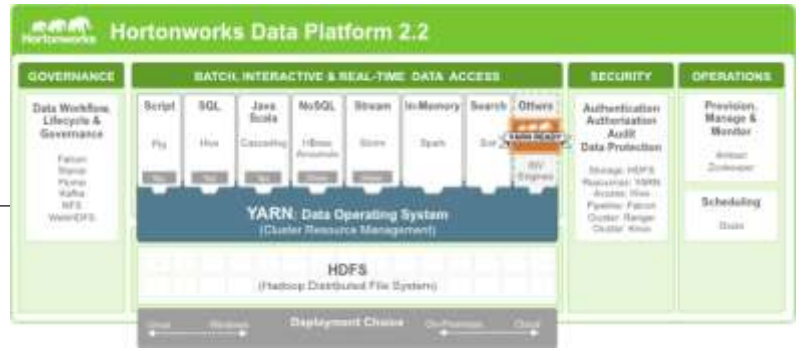
Deploy and effectively manage the platform

Provide deployment choice across physical, virtual, cloud

DEPLOYMENT OPTIONS

Hortonworks

Security needs are changing



Security needs are changing

- YARN unlocks the data lake
- Multi-tenant: Multiple applications for data access
- Changing and complex compliance environment
- Data classification

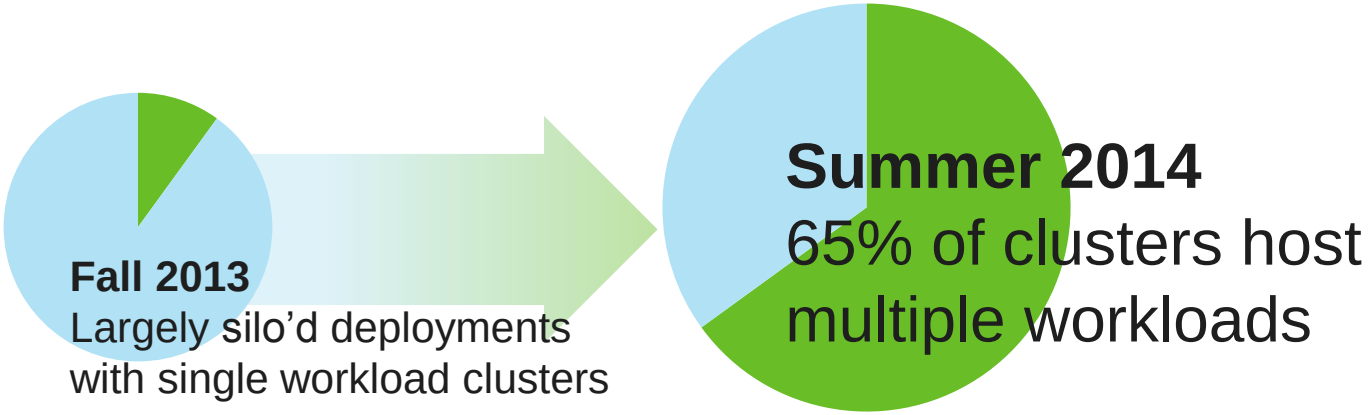
Administration
Centrally management & consistent security

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 today in Hadoop with HDP

Centralized Security Administration

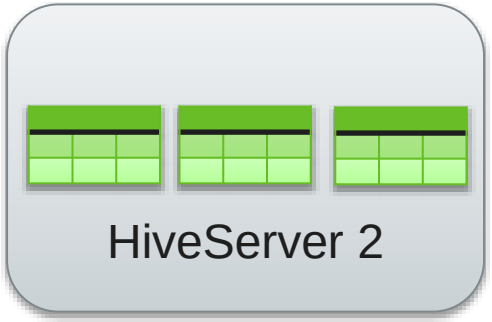
HDP 2.1 + Ranger

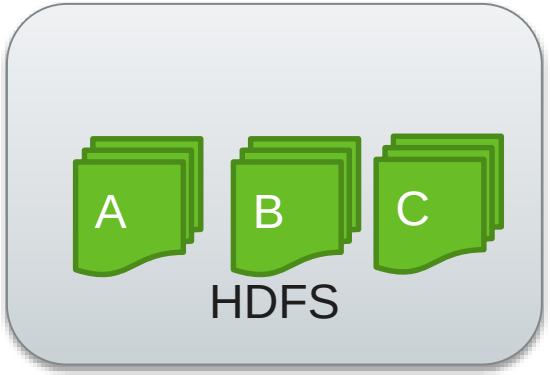
Authentication Who am I/prove it?	**Authorization** Restrict access to explicit data	**Audit** Understand who did what	**Data Protection** Encrypt data at rest & in motion
• Kerberos in native Apache Hadoop • HTTP/REST API Secured with Apache Knox Gateway	• HDFS, Hive and Hbase (Storm and Knox in 2.2) • Fine grain access control	• Centralized audit reporting • Policy and access history	• Wire encryption in Hadoop • Orchestrated encryption with partner tools

Hortonworks

Typical Flow – Hive Access through Beeline client

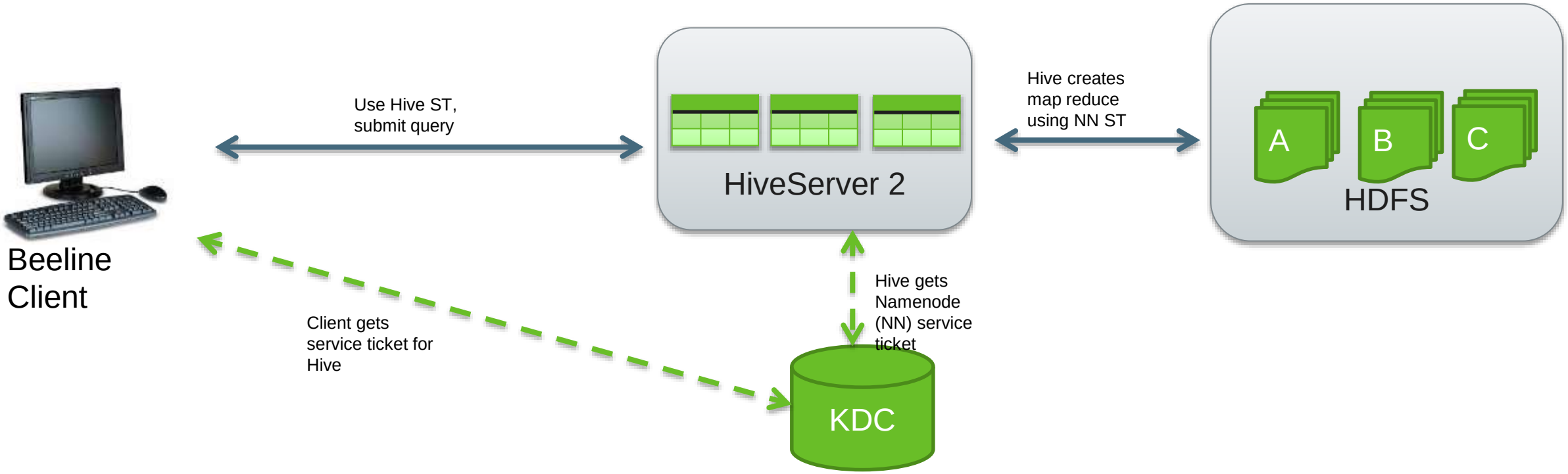
Beeline Client



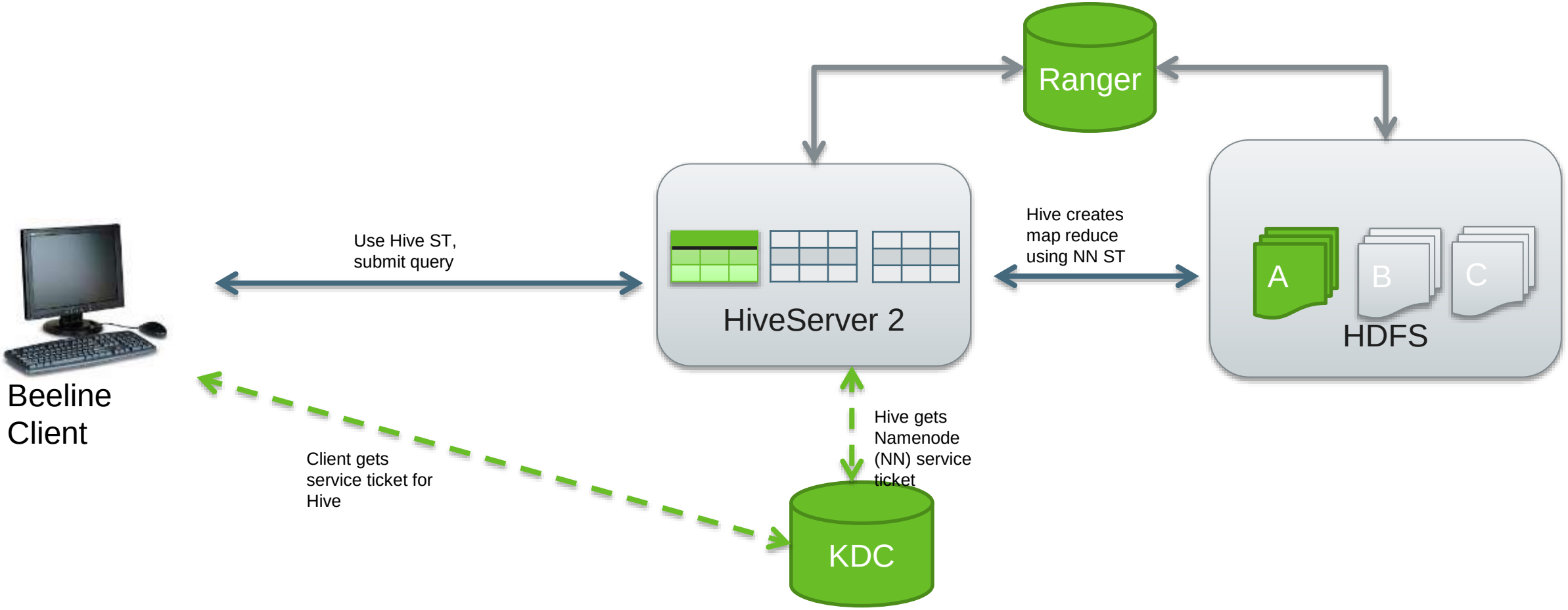


Hortonworks

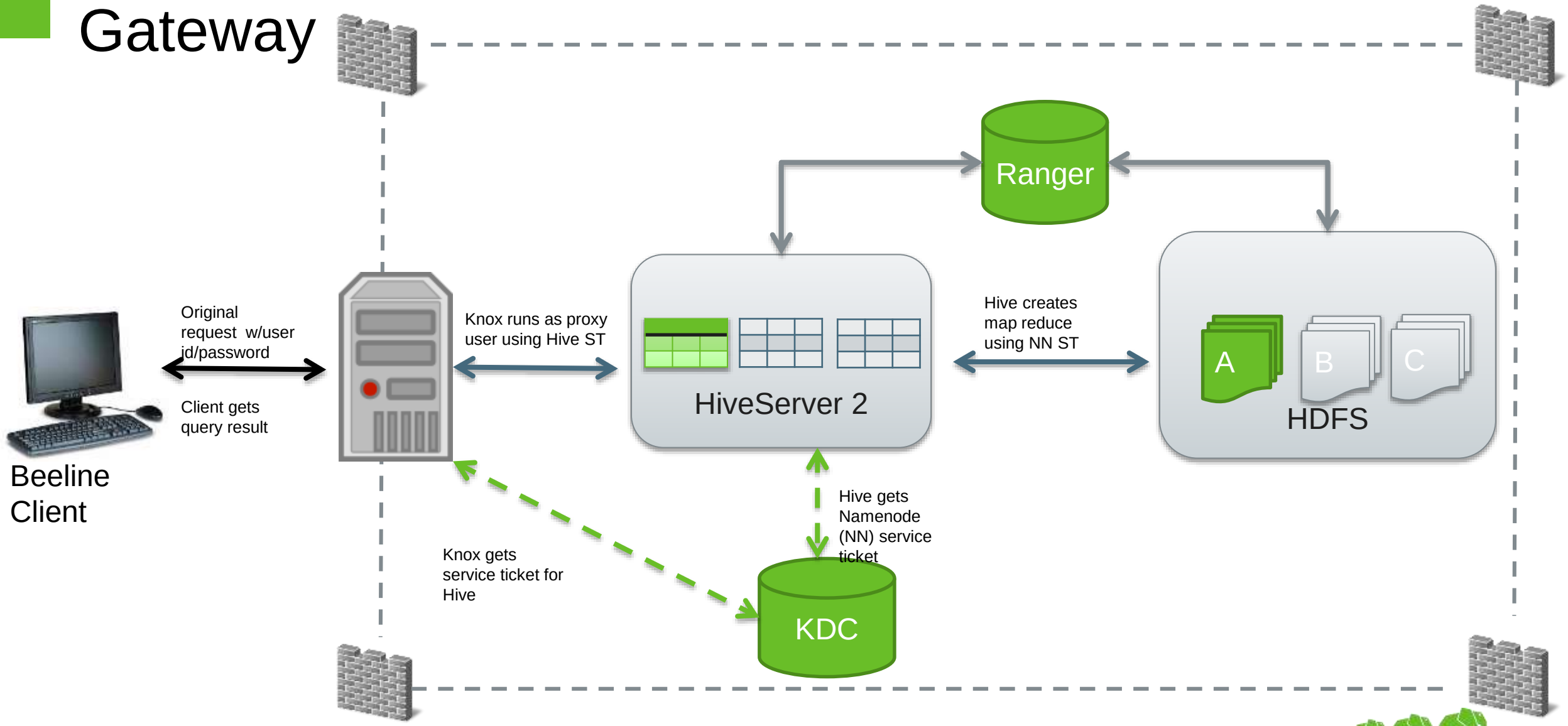
Typical Flow – Authenticate through Kerberos



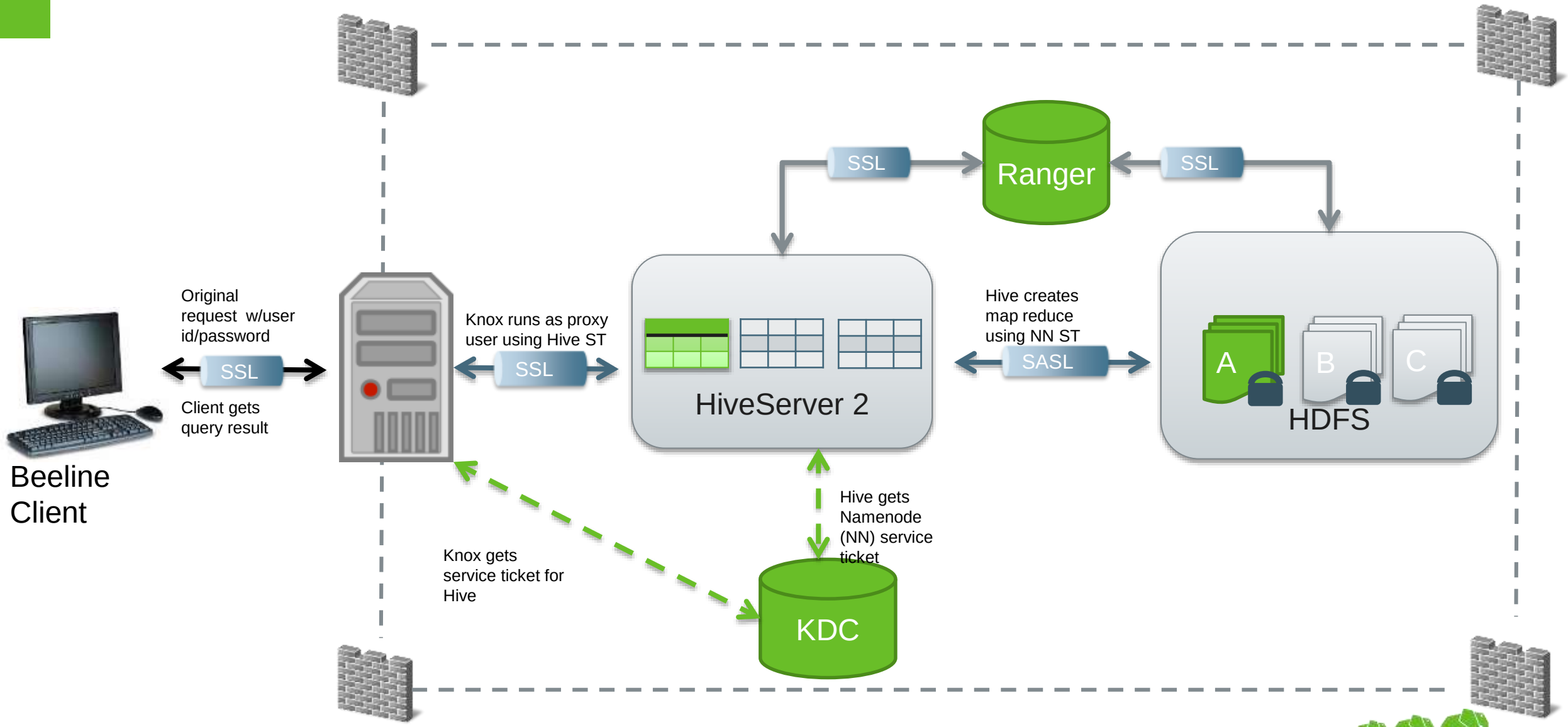
Typical Flow – Add Authorization through Ranger(XA Secure)



Typical Flow – Firewall, Route through Knox Gateway



Typical Flow – Add Wire and File Encryption



Security Features

	HDP Security
Authentication	
Kerberos Support	✓
Perimeter Security – For services and rest API	✓
Authorizations	
Fine grained access control	HDFS, Hbase and Hive, Storm and Knox (next release)
Role base access control	✓
Column level	✓
Permission Support	Create, Drop, Index, lock, user
Auditing	
Resource access auditing	Extensive Auditing
Policy auditing	✓

Hortonworks

Security Features

	HDP Security
Data Protection	
Wire Encryption	✓
Volume Encryption	✓
File/Column Encryption	HDFS TDE & Partners
Reporting	
Global view of policies and audit data	✓
Manage	
User/ Group mapping	✓
Global policy manager, Web UI	✓
Delegated administration	✓

Hortonworks

Authorization and Auditing

Apache Ranger

Hortonworks

Authorization and Audit

Authorization

Fine grain access control

- HDFS – Folder, File
- Hive – Database, Table, Column
- HBase – Table, Column Family, Column

Flexibility in defining policies

Audit

Extensive user access auditing in HDFS, Hive and HBase

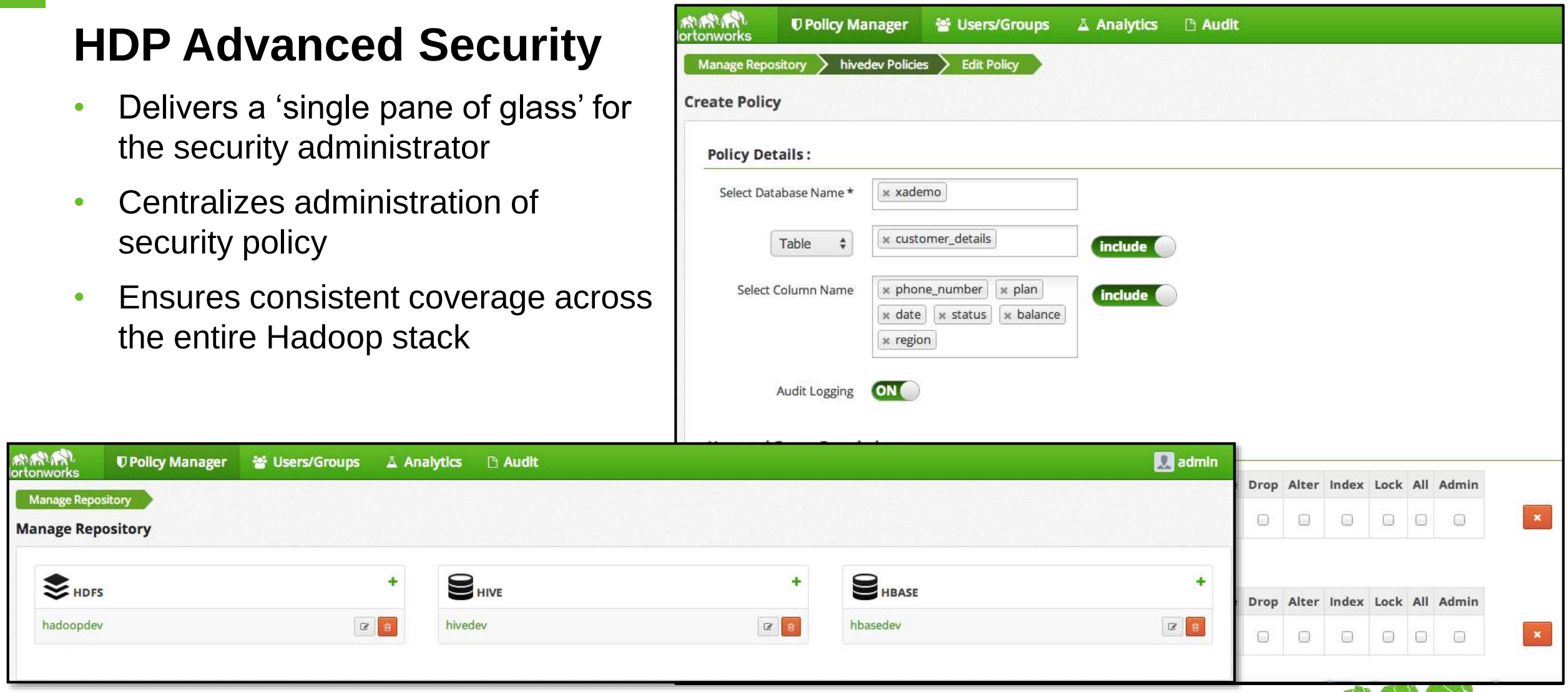
- IP Address
- Resource type/ resource
- Timestamp
- Access granted or denied

Control access into system

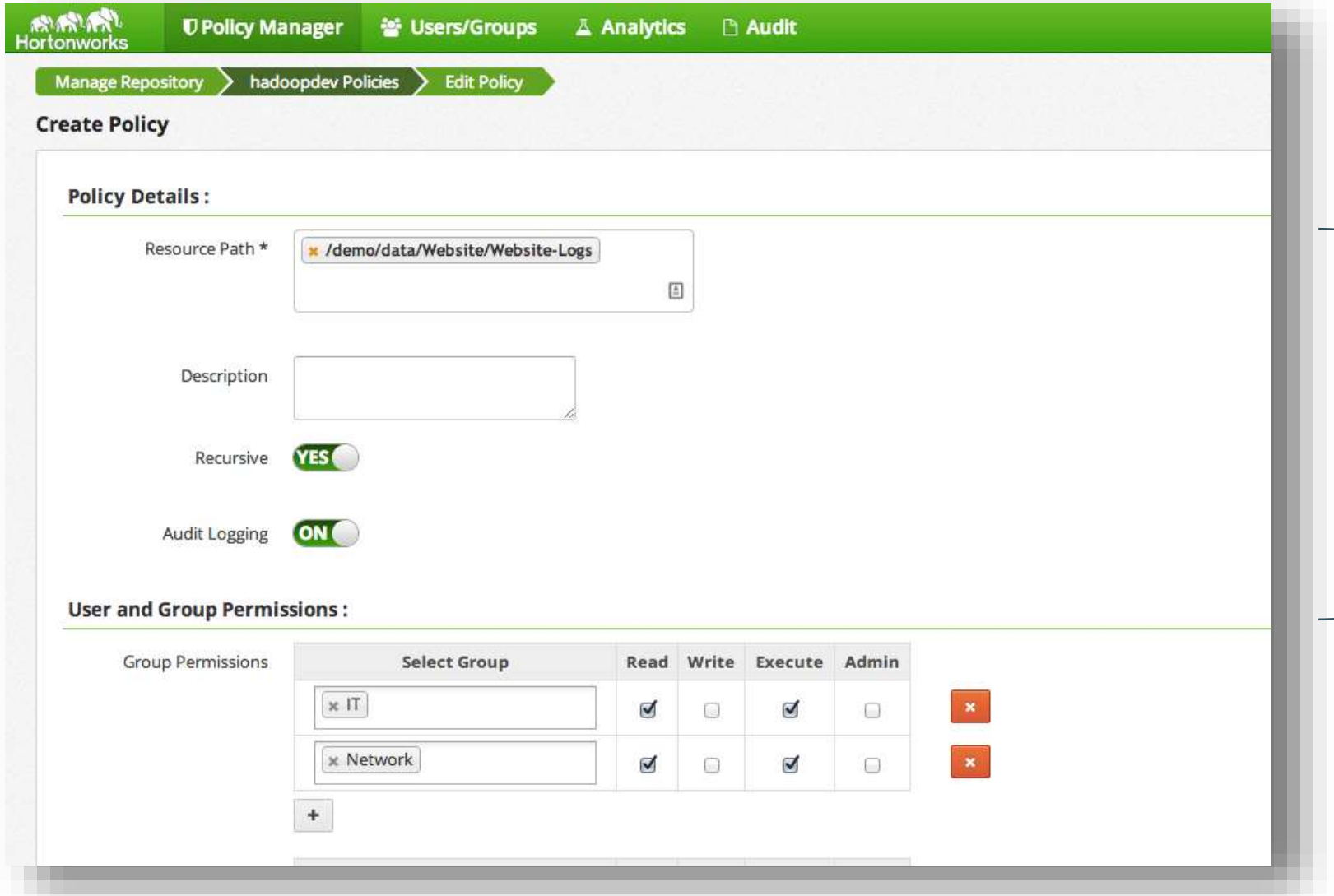
Central Security Administration

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



Setup Authorization Policies



file level access control, flexible definition

Control permissions

Monitor through Auditing

Hortonworks | Policy Manager | Users/Groups | Analytics | Audit | admin

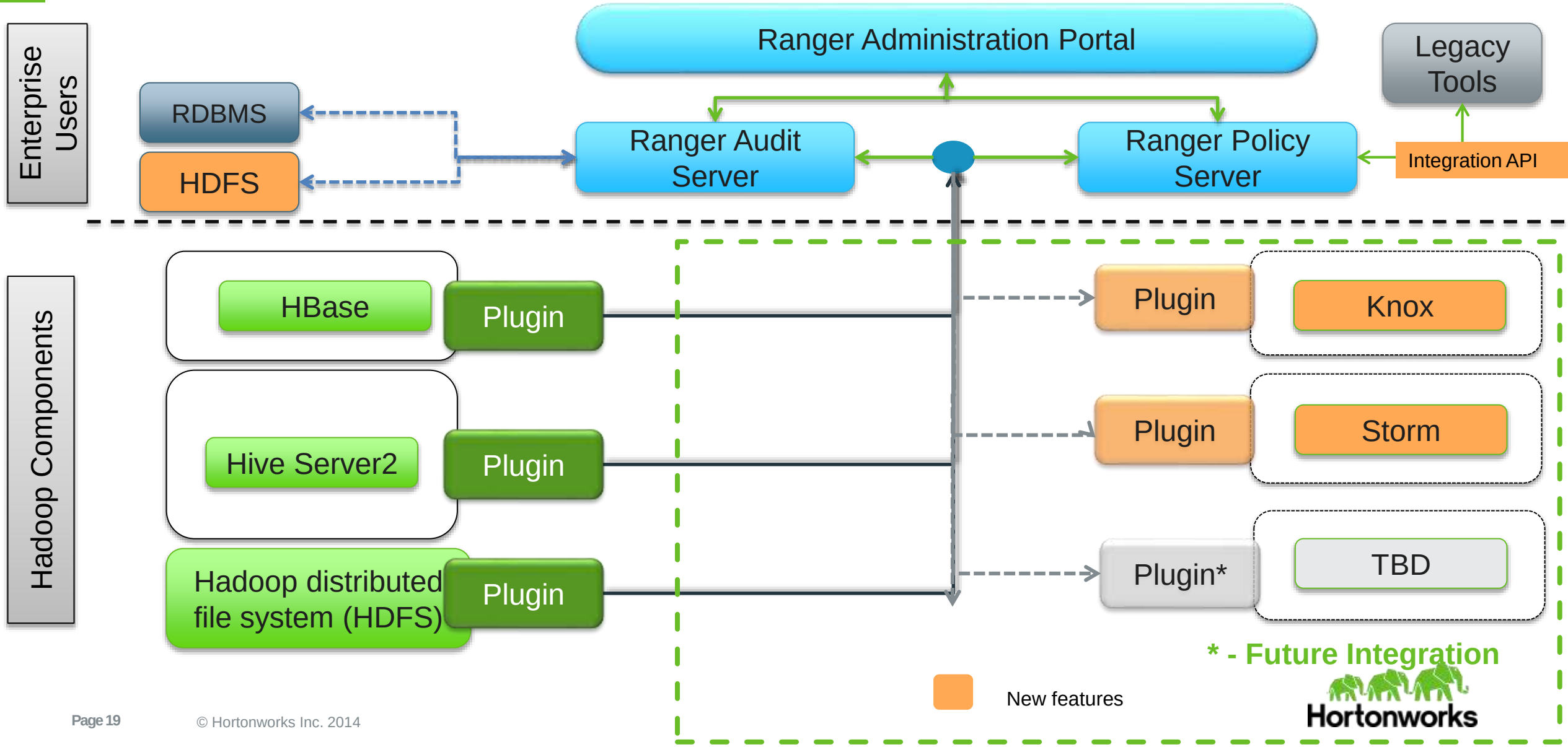
Big Data | Admin | Login Sessions | Agents

START DATE: 06/20/2014

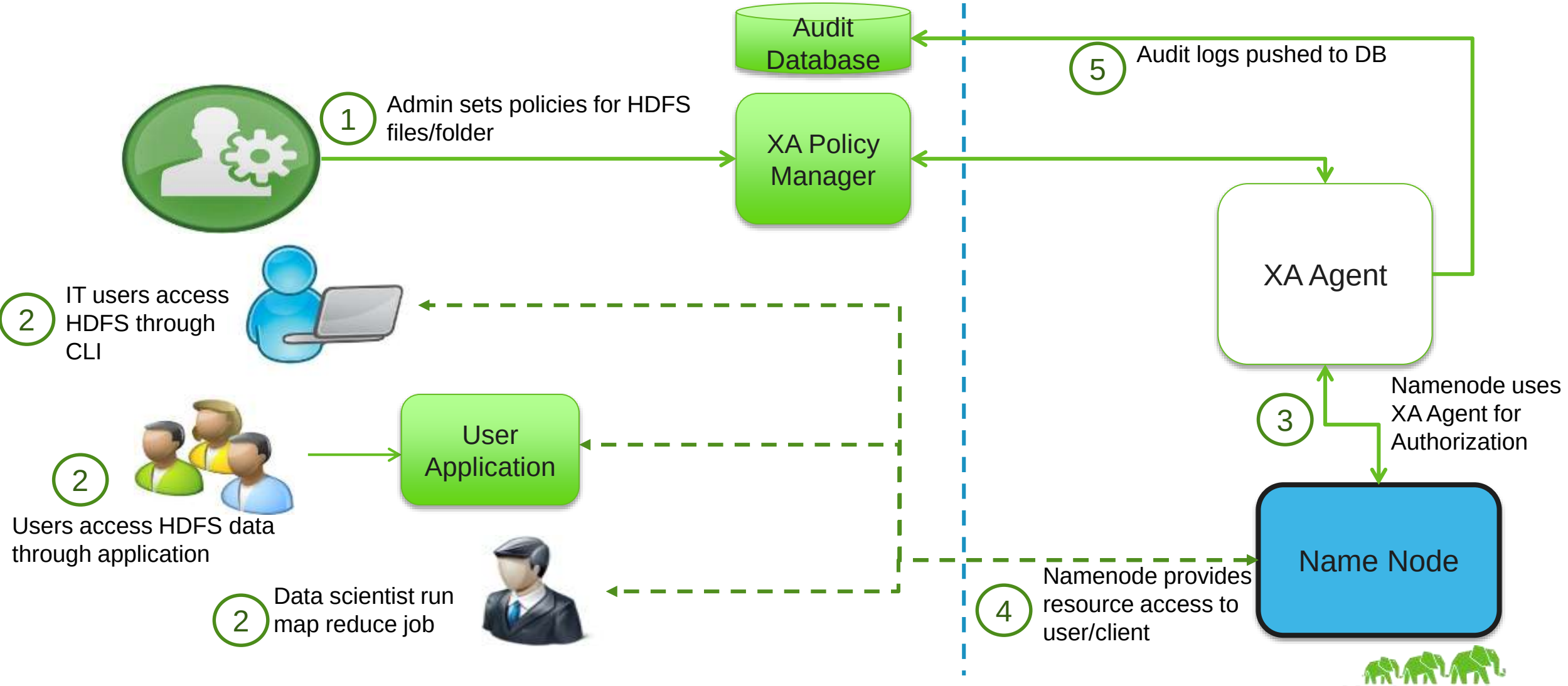
Last Updated Time : 06/20/2014 05:32:12 PM

Event Time	User	Repository Name / Type	Resource Name	Access Type	Result	Access Enforcer	Client IP
06/20/2014 05:31:53 PM	mapred	hadoopdev HDFS	/mr-history/tmp	READ_EXECUTE	Allowed	hadoop-acl	/10.0.2.15
06/20/2014 05:31:30 PM	mapred	hadoopdev HDFS	/mr-history/tmp	READ_EXECUTE	Allowed	hadoop-acl	/10.0.2.15
06/20/2014 05:30:30 PM	mapred	hadoopdev HDFS	/mr-history/tmp	READ_EXECUTE	Allowed	hadoop-acl	/10.0.2.15
06/20/2014 05:29:30 PM	mapred	hadoopdev HDFS	/mr-history/tmp	READ_EXECUTE	Allowed	hadoop-acl	/10.0.2.15
06/20/2014 05:28:53 PM	mapred	hadoopdev HDFS	/mr-history/tmp	READ_EXECUTE	Allowed	hadoop-acl	/10.0.2.15
06/20/2014 05:28:30 PM	mapred	hadoopdev HDFS	/mr-history/tmp	READ_EXECUTE	Allowed	hadoop-acl	/10.0.2.15
06/20/2014 05:27:30 PM	mapred	hadoopdev HDFS	/mr-history/tmp	READ_EXECUTE	Allowed	hadoop-acl	/10.0.2.15
06/20/2014 05:26:31 PM	mapred	hadoopdev HDFS	/mr-history/tmp	READ_EXECUTE	Allowed	hadoop-acl	/10.0.2.15

Authorization and Auditing w/ Ranger



Simplified Workflow - HDFS



Ranger Investments for HDP 2.2

- **New Components Coverage**
  - Storm Authorization & Auditing
  - Knox Authorization & Auditing
- **Deeper Integration with HDP**
  - Windows Support
  - Integration with Hive Auth API, support grant/revoke commands
  - Support grant/revoke commands in Hbase
- **Enterprise Readiness**
  - Rest APIs for policy manager
  - Store Audit logs locally in HDFS
  - Support Oracle DB
  - Ambari support, as part of Ambari 2.0 release

REST API Security through Knox

Securely share Hadoop Cluster

Hortonworks

Hadoop REST API with Knox

Service	Direct URL	Knox URL
WebHDFS	http://namenode-host:50070/webhdfs	https://knox-host:8443/webhdfs
WebHCat	http://webhcat-host:50111/templeton	https://knox-host:8443/templeton
Oozie	http://ooziehost:11000/oozie	https://knox-host:8443/oozie
HBase	http://hbasehost:60080	https://knox-host:8443/hbase
Hive	http://hivehost:10001/cliservice	https://knox-host:8443/hive
YARN	http://yarn-host:yarn-port/ws	https://knox-host:8443/resourcemanager

Masters could be on many different hosts

One hosts, one port

SSL config at one host

Consistent paths

Why Knox?

Enhanced Security

- Protect network details
- SSL for non-SSL services
- WebApp vulnerability filter

Centralized Control

- Central REST API auditing
- Service-level authorization
- Alternative to SSH “edge node”

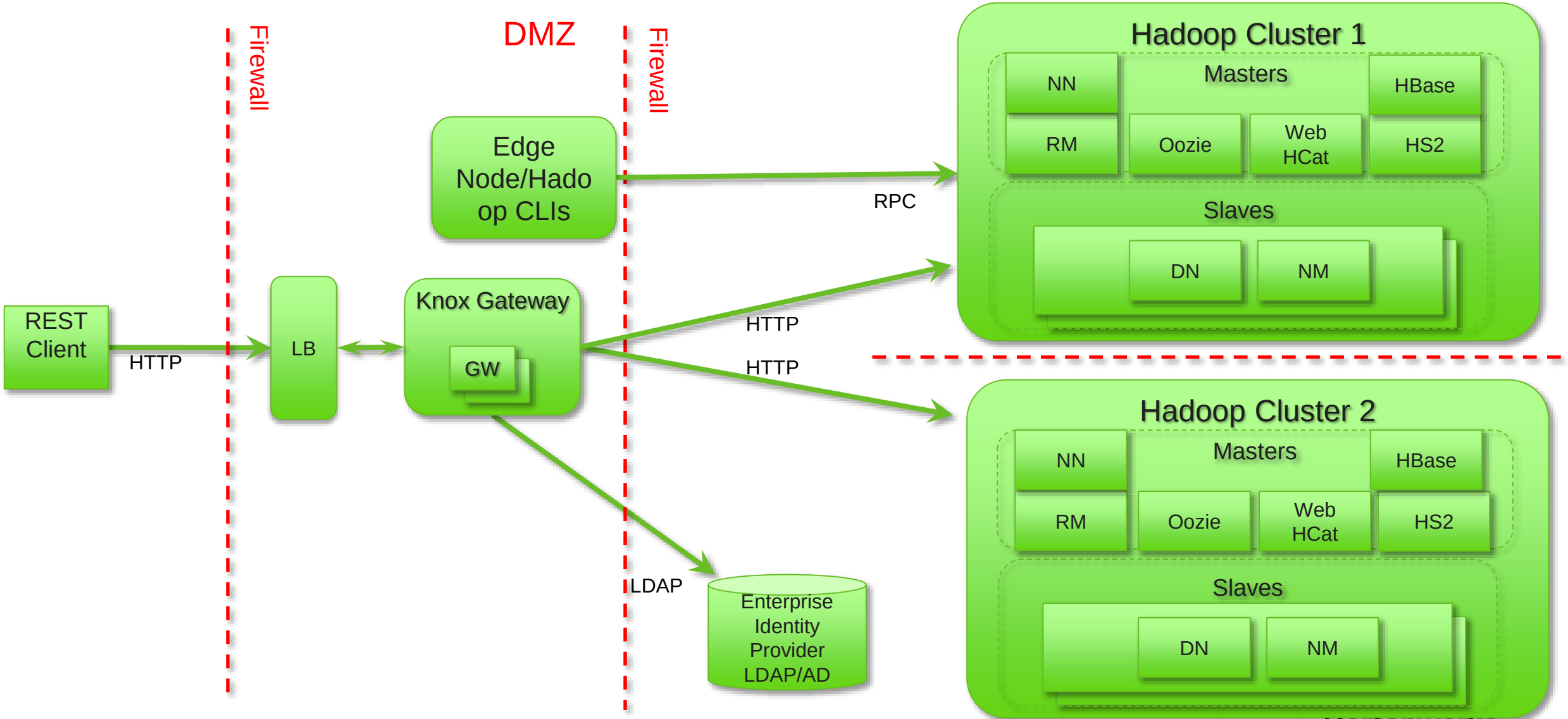
Simplified Access

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

Enterprise Integration

- LDAP integration
- Active Directory integration
- SSO integration
- Apache Shiro extensibility
- Custom extensibility

Hadoop REST API Security: Drill-Down



What’s New in Knox with HDP 2.2

- Use Ambari for Install/start/stop/configuration
- Knox support for HDFS HA
- Support for YARN REST API
- Support for SSL to Hadoop Cluster Services (WebHDFS, HBase, Hive & Oozie)
- Knox Management REST API
- Integration with Ranger (fka XA Secure) to for Knox Service Level Authorization

Workshop: Enabling Security

Hortonworks

Let’s Begin

- We will use HDP Sandbox with FreeIPA Software Installed
  - FreeIPA is an integrated security information management solution combining Linux (Fedora), 389 Directory Server, MIT Kerberos, NTP, DNS, Dogtag (Certificate System). It consists of a web interface and command-line administration tools
  - In the workshop we use FreeIPA for User Identity Management
  - **Note**: Steps outlined in the workshop are applicable for other identity management solutions such as Active Directory

Authentication

1. **Create end users and groups in FreeIPA**
   – End Users will query HDP via Hue, Beeline & JDBC/ODBC clients

2. **Enable Kerberos for the HDP Cluster**
   – Hadoop now authenticates all access to the cluster

3. **Integrate Hue with FreeIPA**
   – Users are validated against FreeIPA

4. **Configure Linux to use FreeIPA as central store of posix data using nslcd**
   – Enables Hadoop to determine user groups without requiring a local linux user account

We have now set Authentication

– A user can open a shell, authenticate using kinit and submit hadoop commands or alternatively log into HUE to access Hadoop.

Enable Perimeter Security

1. **KNOX Is Available on Sandbox**
   - Enables Perimeter Security. Enables single point of cluster access using Hadoop REST APIs, JDBC and ODBC calls

2. **Configure KNOX to authenticate against FreeIPA**

3. **Configure WebHDFS & Hiveserver2 to support JDBC/ODBC access over HTTP**

4. **Use Excel to access Hive via KNOX**
   - Note, Knox eliminates the need to secure Kerberos ticket on the client machine for user authentication

We have now set Perimeter Security

- Users can now access the cluster via the Gateway services

Authorization & Audit

1. **Install Apache Ranger**
   – Comprehensive authorization and audit tool for Hadoop
2. **Sync users between Apache Ranger and FreeIPA**
   – Note, end users are only required to be maintained in **one** enterprise identity management system
3. **Configure HDFS & Hive to use Apache Ranger**
   – In this workshop we will only show steps as it relates to hive authorization. Similar capabilities are available for other HDP components.
4. **Define HDFS & Hive Access Policy For Users**
   – User “hive” is a special user and must be assigned universal access
5. **Log into Hue as the end user and note the authorization policies being enforced**
   – Review Audit Information

We have now set Authorization & Audit

– All user access to a Hive is governed & audited by policies maintained in Apache Ranger.

Encryption

1. **Wire Level Encryption**
   - Follow instruction here http://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.0.6.0/bk_reference/content/ch_wire6.html

2. **Volume Level Encryption**
   - Leverage LUKS. Sample script provided

3. **Column level encryption & data masking**
   - Collaborate with our key security partners

Resources

Hortonworks

Security Page

Apache Ranger

hortonworks.com/hadoop/ranger/

Hortonworks

WHAT IS HADOOP? » APACHE RANGER

Apache Ranger

Comprehensive security for Enterprise Had

Ranger delivers a comprehensive approach to
central security policy administration across th
authentication, authorization, accounting and

It already extends baseline features for coordinated enforcem
SQL and real–time IN Hadoop. And we plan to leverage the
policies consistently against additional Hadoop ecosystem c
Storm, Solr, Spark, and more. It truly represents a major ste
comprehensive approach to security – all completely as ope

Hortonworks Investment Themes for Ranger

Investment Theme	Planned Enhancements
Extension of support	Additional investments extend administration of authorization and auditing to more

Resources

Discover HDP2.1: Apache Storm for Stream Data

Apache Knox Gateway

hortonworks.com/hadoop/knox-gateway/

Hortonworks

ENGLISH Login Register Search Go

ENTERPRISE HADOOP SOLUTIONS PRODUCTS LABS TRAINING GET STARTED PARTNERS ABOUT BLOG

WHAT IS HADOOP? » APACHE KNOX GATEWAY

Apache Knox Gateway

A single point of secure access for Hadoop clusters

The Knox Gateway ("Knox") is a system that provides a single point of authentication and access for Apache™ Hadoop® services in a cluster. The goal of the project is to simplify Hadoop security for users who access the cluster data and execute jobs, and for operators who control access and manage the cluster. Knox runs as a server (or cluster of servers) that serve one or more Hadoop clusters.

What Knox Gateway Does

Knox Gateways provides security for multiple Hadoop clusters, with these advantages:

- **Provide perimeter security** to make Hadoop security setup easier
- **Support authentication and token verification** security scenarios
- **Deliver users a single cluster end-point** that aggregates capabilities for data and jobs
- **Enable integration** with enterprise and cloud identity management environments

KNOX

February 2013

12

Hortonworks Sandbox

Try Knox Gateway with Sandbox

Hortonworks Sandbox is a self-

Contact Us

Hortonworks

Hortonworks Security Investment Plans

HDP + XA

Comprehensive Security for Enterprise Hadoop

Goals:

Comprehensive Security

Meet all security requirements across Authentication, Authorization, Audit & Data Protection for all HDP components

Central Administration

Provide one location for administering security policies and audit reporting for entire platform

Consistent Integration

Integrate with other security & identity management systems, for compliance with IT policies

…all **IN** Hadoop

Investment themes

Previous Phases — Delivered

- ✓ Kerberos Authentication
- ✓ HDFS, Hive & Hbase authorization
- ✓ Wire Encryption for data in motion
- ✓ Knox for perimeter security
- ✓ Basic Audit in HDFS & MR
- ✓ SQL Style Hive Authorization
- ✓ ACLs for HDFS

XA Secure Phase — Delivered XA Secure

- Centralized Security Admin for HDFS, Hive & HBase
- Centralized Audit Reporting
- Delegated Policy Administration

Future Phases

- Encryption in HDFS, Hive & Hbase
- Centralized security administration of entire Hadoop platform
- Centralized auditing of entire platform
- Expand Authentication & SSO integration choices
- Tag based global policies (e.g. Policy for PII)

Q&A

Hortonworks