



SECURING DATA IN HYBRID ENVIRONMENTS USING APACHE RANGER

Don Bosco Durai, Privacera

Apache Ranger PMC

Madhan Neethiraj, Cloudera

Apache Ranger PMC, Apache Atlas PMC

DISCLAIMER

- *This document may contain product features and technology directions that are under development, may be under development in the future or may ultimately not be developed.*
- *Project capabilities are based on information that is publicly available within the Apache Software Foundation project websites ("Apache"). Progress of the project capabilities can be tracked from inception to release through Apache, however, technical feasibility, market demand, user feedback and the overarching Apache Software Foundation community development process can all effect timing and final delivery.*
- *This document's description of these features and technology directions does not represent a contractual commitment, promise or obligation from Cloudera and Privacera to deliver these features in any generally available product.*
- *Product features and technology directions are subject to change, and must not be included in contracts, purchase orders, or sales agreements of any kind.*
- *Since this document contains an outline of general product development plans, customers should not rely upon it when making purchasing decisions.*

ABOUT PRIVACERA

CLOUD ACCESS MANAGER

CLOUD
DISCOVERY

CLOUD
ANONYMIZATION

Storage



SQL



No SQL



Streaming,
Serverless,
ML



AGENDA

Apache Ranger overview

Security Challenges Hybrid Deployment

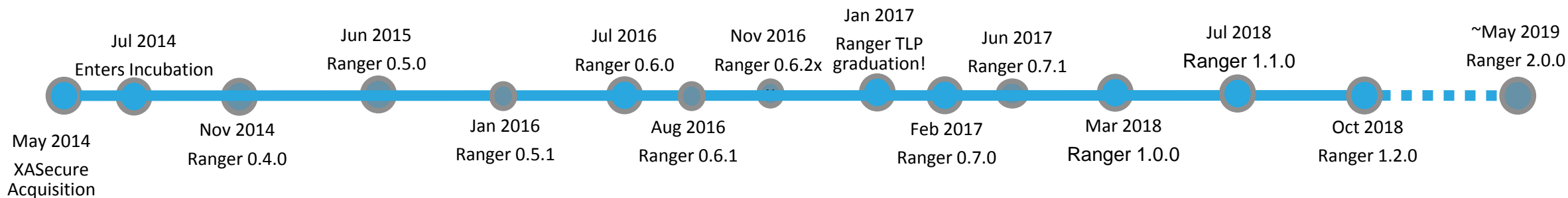
Implementing Hybrid Security using Ranger

New Features: Security Zones, Role Based Access Control, Conditions at Policy Scope

Demo

Questions

APACHE RANGER: OVERVIEW - HISTORY



Ranger 0.7.x

Tag based Masking
Export/import of Policies
\$User and macros
User Sync Nested LDAP Support
Plugin status tab
“Show columns” and
“describe extended support”
Incremental LDAP Sync

Ranger 1.1.0

Time based policies
Metadata security
Audit only (compliance) role
Hive UDF usage authorization
Show Hive query in audits
Policy labels
Audit enhancements

Ranger 2.0.0

Hadoop3 version updates
Security zones
Policy level custom conditions
Role based authorization
DB Schema optimization for faster policy CRUD
Hadoop Trusted-proxy authentication

- Committers: 29
- Contributors from:
eBay, MSFT,
Huawei, Pandora,
Accenture, ING,
Talend, ZTE

APACHE RANGER: OVERVIEW – FEATURES

- Centralized policy administration
- Centralized auditing
- Dynamic row filtering
- Dynamic data masking
- Tag based authorization and data-masking policies
- Rich & extendable policy enforcement engine
- Key Management System (KMS)
- **New Feature: Security Zones**
- **New Feature: Support for Roles Based Access Control**
- **New Feature: Conditions at policy scope**

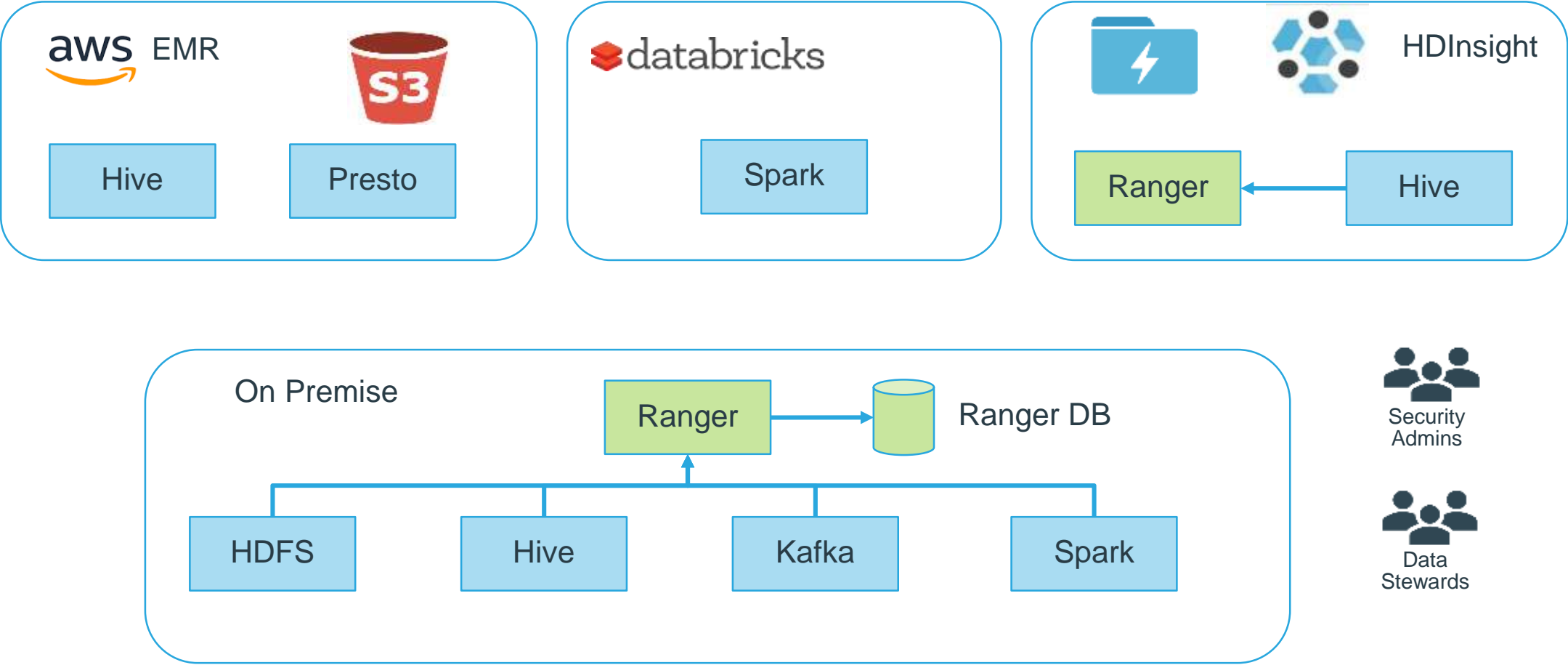
APACHE RANGER: OVERVIEW – CENTRALIZED AUTHORIZATION

The screenshot displays the Apache Ranger web interface. At the top, a green navigation bar contains the 'Ranger' logo, 'Access Manager', 'Audit', and 'Settings' links, along with a user profile icon labeled 'admin'. Below this, a 'Service Manager' tab is selected. The main content area is titled 'Service Manager' and includes 'Import' and 'Export' buttons. It features a grid of service configuration cards for HDFS, HBASE, HIVE, YARN, KNOX, STORM, SOLR, KAFKA, NIFI, NIFI-REGISTRY, and ATLAS. Each card shows a folder icon, the service name, a list of configured services (e.g., 'brown_hadoop' for HDFS), and action icons for adding, editing, and deleting.

Service	Configured Services
HDFS	brown_hadoop
HBASE	brown_hbase
HIVE	brown_hive
YARN	brown_yarn
KNOX	brown_knox
STORM	
SOLR	
KAFKA	brown_kafka
NIFI	
NIFI-REGISTRY	
ATLAS	brown_atlas

SECURITY IN HYBRID ENVIRONMENT

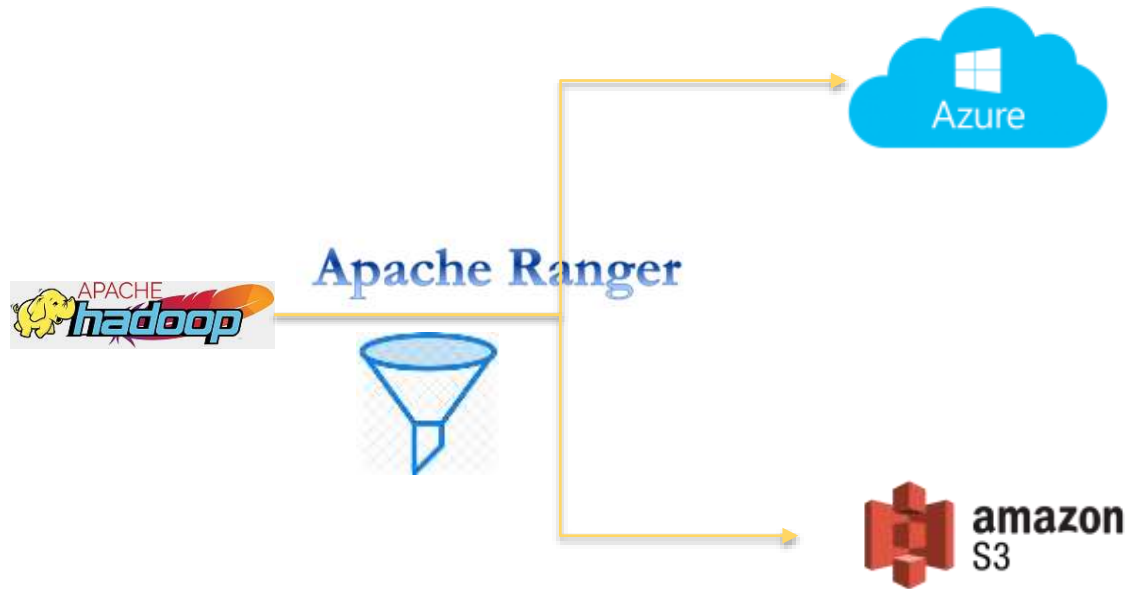
HYBRID DEPLOYMENT: OVERVIEW



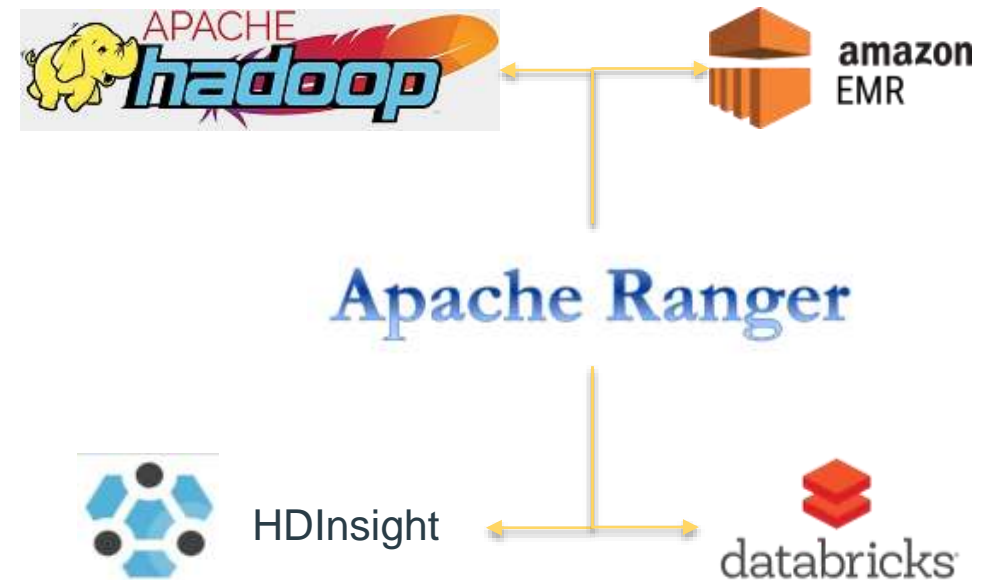
HYBRID DEPLOYMENT: SECURITY CHALLENGES

- Every environment has different security model
- Access policies needs to be set in each environment
- Policies needs to be consistent
- The granularity of access control are not the same
- Policies can go out of sync very soon
- Regulation and compliance requirements on what data can be copied to cloud and whether it should be encrypted or deidentified

Option #1 Restrict Data from On-premise



Option #2 Centralized Ranger



HYBRID DEPLOYMENT: OPTION #1

- Filter & Redact data copied to cloud
- Use Hive to export data to S3
- Apply Ranger Row Level Filtering and Column Masking on ETL user (e.g. s3etl)
- Setup cloud native access policies for copied data

APACHE RANGER: ROW-FILTER, COLUMN-MASKING POLICIES

ID	CONSENT	TAX_ID	NAME	EMAIL
1	Y	123456789	John	john@acme.com
2	Y	987654321	Jane	jane@acme.com
3	N	789654123	Mary	mary@acme.com
4	Y	321789654	David	david@acme.com
5	N	456321789	Max	max@acme.com

ID	CONSENT	TAX_ID	NAME	EMAIL
1	Y	XXXXXXXXXX	John	dkrx@acme.com
2	Y	XXXXXXXXXX	Jane	yafe@acme.com
4	Y	XXXXXXXXXX	David	aumd2@acme.com

APACHE RANGER: ROW-FILTER, COLUMN-MASKING POLICIES

Hive Table *

× customer_detail

Select User	Access Types	Row Level Filter
<div>× s3etl</div>	<div>select</div> <div></div>	<div>consent = 'Y'</div> <div></div>

Select User	Access Types	Select Masking Option
<div>× s3etl</div>	<div>select</div> <div></div>	<div>Redact</div> <div></div>

Select User	Access Types	Select Masking Option
<div>× s3etl</div>	<div>select</div> <div></div>	<div>Custom</div> <div></div> <div>privacera.protect('FPE_EMAIL')</div>

HYBRID DEPLOYMENT: OPTION #1 – PROS AND CONS

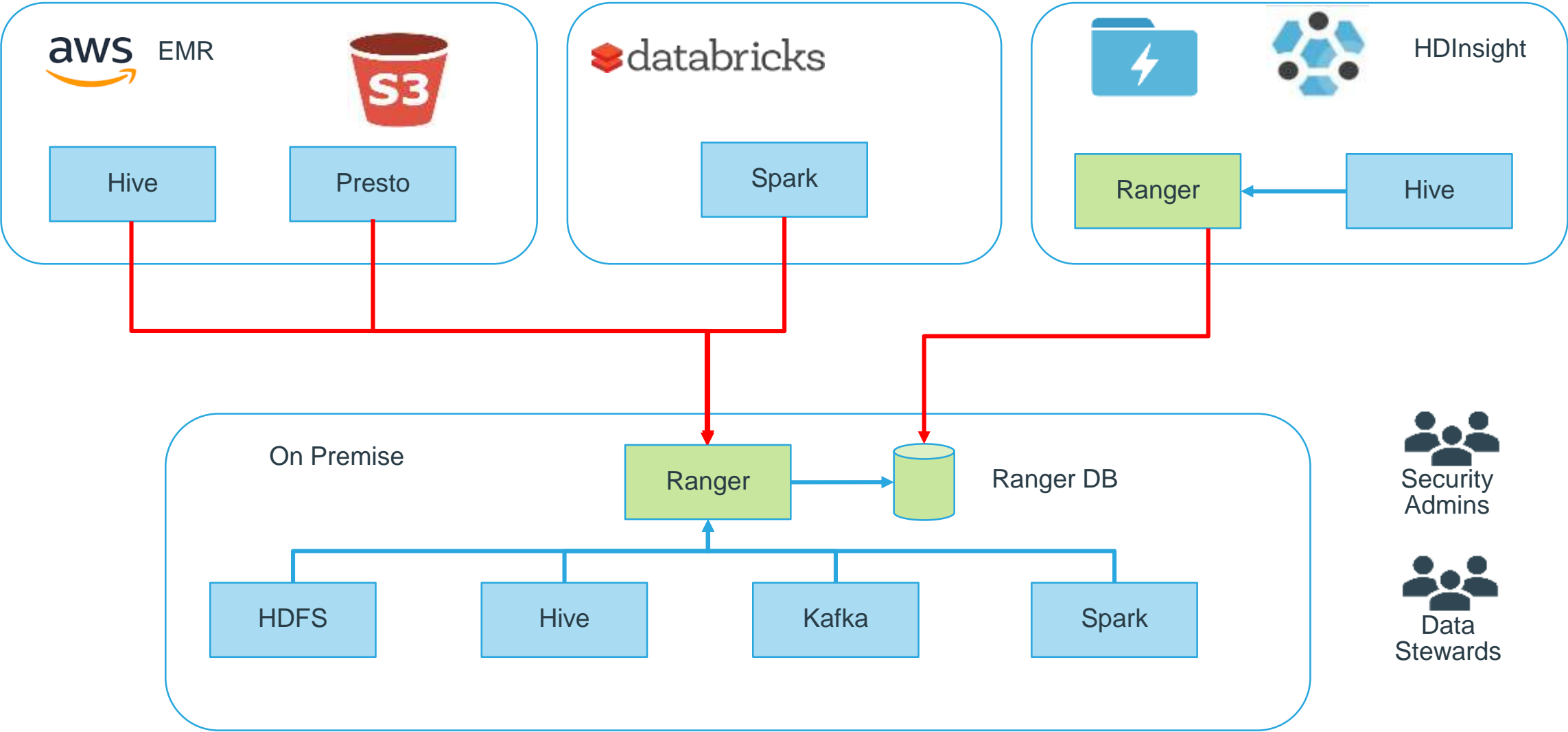
- Advantages

- Simple to implement
- Fine grained policies enforced on premise using Filtering, Redaction and Transformation
- Use cloud security policy for coarse grain policies
- Make data accessible to non-Ranger supported services like AWS Redshift, AWS Athena, SageMaker, etc.

- Limitation

- Not real-time
- If policies changes, then data need to be recopied to cloud
- Need to manage policies on both the sides

HYBRID DEPLOYMENT: OPTION #2 - CENTRALIZED SECURITY



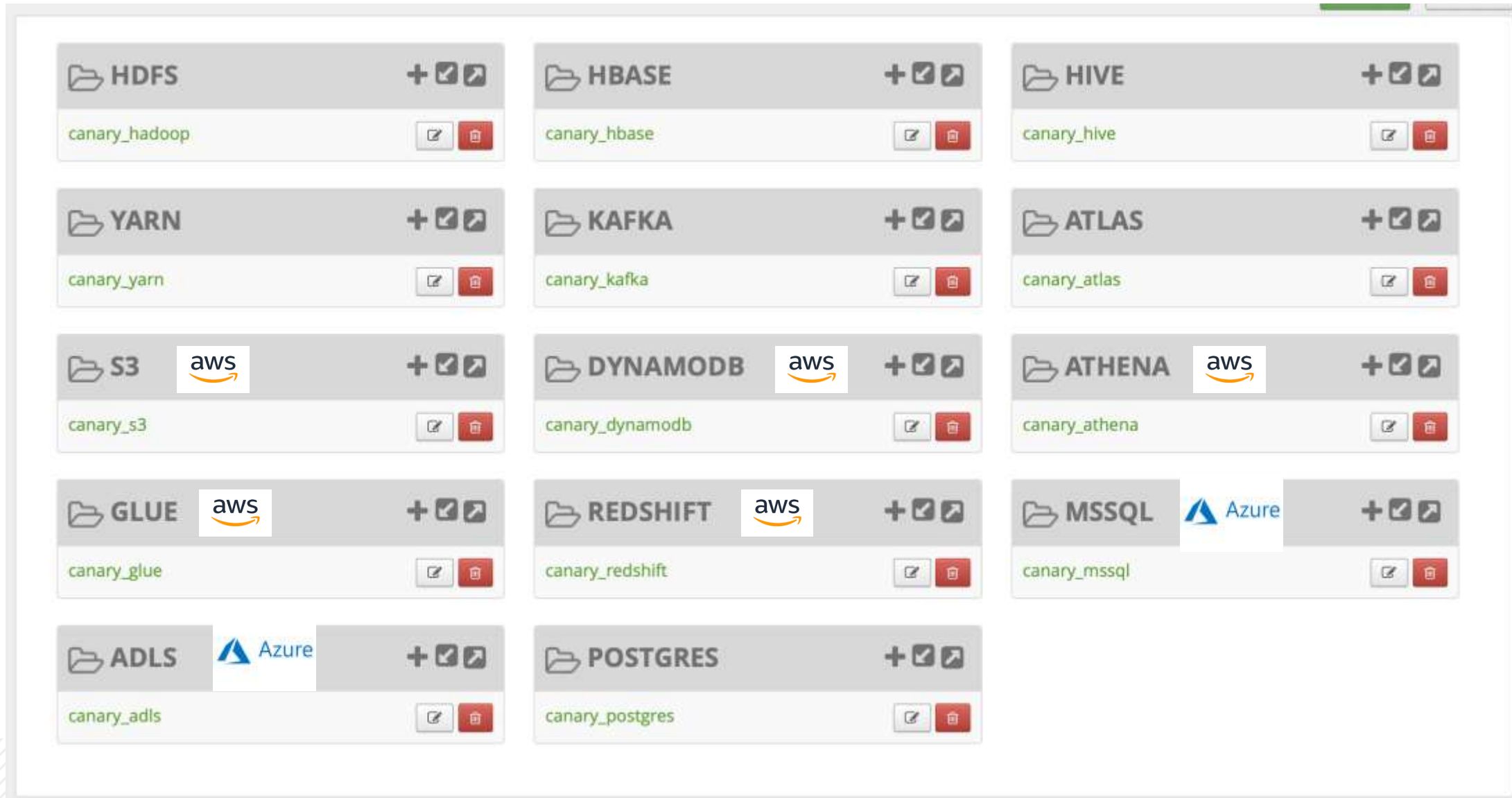
HYBRID DEPLOYMENT: OPTION #2

- Common Ranger Admin or Ranger Database for all environments
- Single Ranger to manage the policies for all environments
- If you are using the same name for resources, e.g. Database, Table and Column name, then a same policy would be used by all the environments
- Tag-based policies can be used to authorize access to cloud-specific data as well
- Use new Ranger features under development to support central policy management
 - Security Zone
 - Scoped Policy
 - Roles in Ranger

HYBRID DEPLOYMENT: OPTION #2 – PROS AND CONS

- **Advantages**
 - Centrally Manage security policies for all environments
 - Policy changes applied in real-time in all environments
 - Leverage Tag Based policies for consistent behavior
 - Increasing support for Ranger by 3rd party vendors. Privacera, StarBurst, Dremio, Microsoft, EMC Isilon, etc.
- **Limitation**
 - Need reliable and secure network connectivity between premise and cloud (site to site VPN)
 - All cloud components might be not supported by Open Source Ranger.
 - Ranger integration for cloud environment is not supported by the community and will require additional setup in the cloud services/deployments

PRIVACERA EXTENSION TO APACHE RANGER



DEMO

SECURITY ZONES

APACHE RANGER: SECURITY ZONES - INTRODUCTION

- Partition resources for easier administration of security policies

Zone	HDFS	Hive	HBase	Kafka
landing	/landing/	db=*landing		
staging	/staging/	db=*staging	table=*staging	
marketing	/marketing	db=marketing	table=marketing	topic=mktg_campaign

- Policies in a zone are applied only for resources included in the zone. For example:
 - a *landing* zone policy for *db=** applies only for the resources of *landing* zone. It will not impact other resources, like *db=marketing*
- Policy administration for each zone can be delegated to specific users/groups

APACHE RANGER: SECURITY ZONES - INTRODUCTION

- Audit log includes zone name, allows to quickly filter accesses to resources of a zone
- REST API for Security Zone administration
- Example use cases:
 - ‘on-prem’ zone for resources that should only be accessible from on-prem clusters
 - ‘test-data’ zone for resources that can be used for test purposes by wider set of users/groups, without impacting production data

APACHE RANGER: SECURITY ZONES - ADMINISTRATION

Security Zone

landing

Edit Delete

Zone Administration

Admin Users

Admin Usergroups

Auditor Users

Auditor Usergroups

Service Name	Service Type	Resource
cl1_hadoop	HDFS	path : /landing
cl1_hive	HIVE	database : *landing

Security Zone

staging

Edit Delete

Zone Administration

Admin Users

Admin Usergroups

Auditor Users

Auditor Usergroups

Service Name	Service Type	Resource
cl1_hive	HIVE	database : *staging
cl1_hbase	HBASE	table : *staging

Security Zone

marketing

Edit Delete

Zone Administration

Admin Users

Admin Usergroups

Auditor Users

Auditor Usergroups

Service Name	Service Type	Resource
cl1_hadoop	HDFS	path : /marketing
cl1_hive	HIVE	database : marketing
cl1_hbase	HBASE	table : marketing
cl1_kafka	KAFKA	topic : mktg_campaign, mktg_events

APACHE RANGER: SECURITY ZONES - ADMINISTRATION

Security Zone

marketing

EditDelete

Zone Administration

Admin Users

Admin Usergroupsmktgadmins

Auditor Users

Auditor Usergroupsmktgadmins

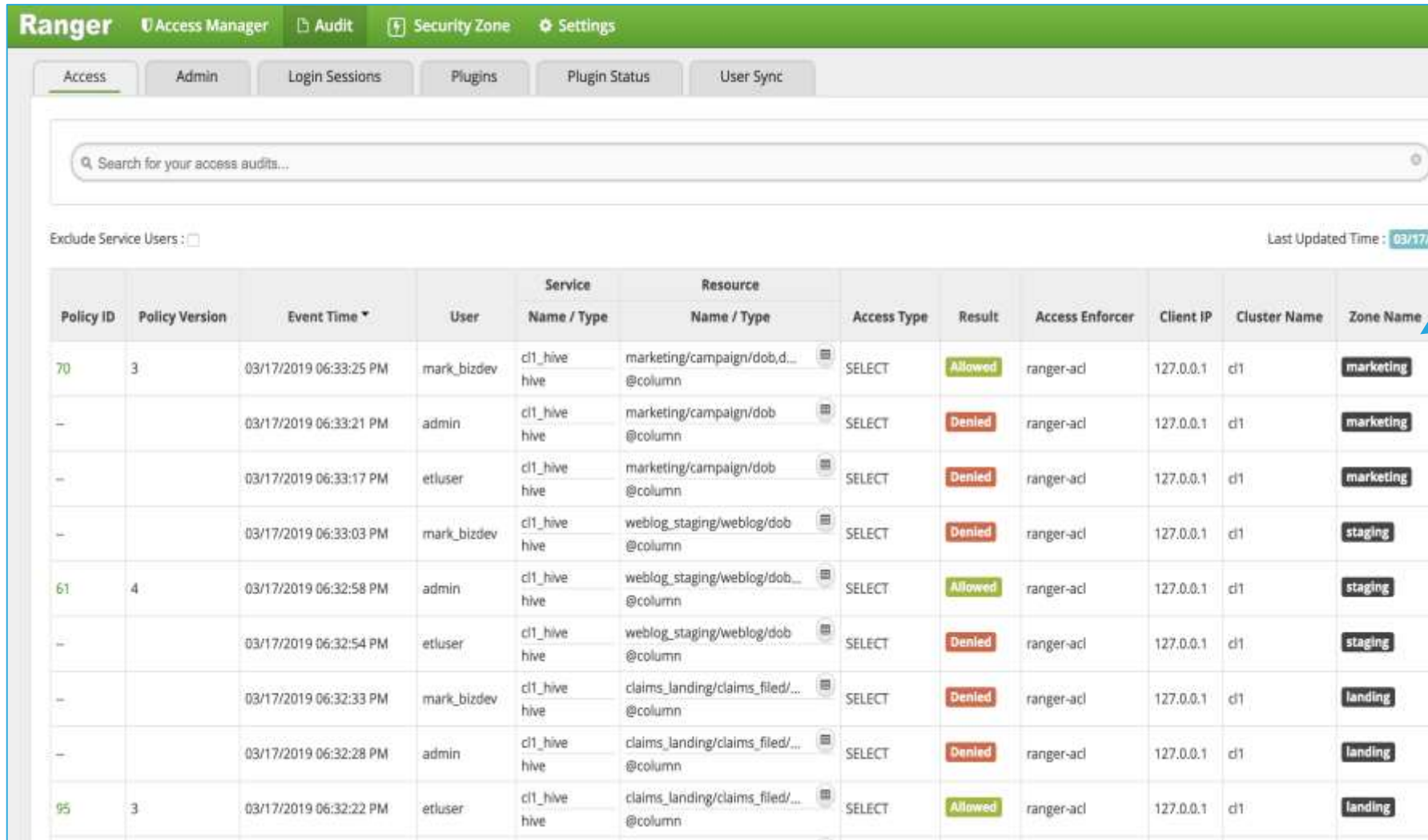
Service Name	Service Type	Resource
cl1_hadoop	HDFS	path : /marketing
cl1_hive	HIVE	database : marketing
cl1_hbase	HBASE	table : marketing
cl1_kafka	KAFKA	topic : mktg_campaign, mktg_events

APACHE RANGER: SECURITY ZONES - POLICY ADMINISTRATION

The screenshot displays the Apache Ranger web interface for policy administration. The top navigation bar includes 'Ranger', 'Access Manager', 'Audit', 'Security Zone', and 'Settings'. The user 'admin' is logged in. The 'Service Manager' section shows a grid of services. A dropdown menu for 'Security Zone' is open, showing a list of zones: 'landing', 'staging', 'marketing', and 'hybrid'. The 'landing' zone is selected and highlighted. A blue arrow points from the 'landing' zone to the list of bullet points on the right.

- Users see only zones in which they have admin privileges
- Zone support extends to access, data-masking, row-filter and tag-based policies

APACHE RANGER: SECURITY ZONES – AUDIT LOGS



Policy ID	Policy Version	Event Time	User	Service Name / Type	Resource Name / Type	Access Type	Result	Access Enforcer	Client IP	Cluster Name	Zone Name
70	3	03/17/2019 06:33:25 PM	mark_bizdev	cl1_hive hive	marketing/campaign/dob,d...@column	SELECT	Allowed	ranger-acl	127.0.0.1	d1	marketing
—		03/17/2019 06:33:21 PM	admin	cl1_hive hive	marketing/campaign/dob...@column	SELECT	Denied	ranger-acl	127.0.0.1	d1	marketing
—		03/17/2019 06:33:17 PM	etluser	cl1_hive hive	marketing/campaign/dob...@column	SELECT	Denied	ranger-acl	127.0.0.1	d1	marketing
—		03/17/2019 06:33:03 PM	mark_bizdev	cl1_hive hive	weblog_staging/weblog/dob...@column	SELECT	Denied	ranger-acl	127.0.0.1	d1	staging
61	4	03/17/2019 06:32:58 PM	admin	cl1_hive hive	weblog_staging/weblog/dob...@column	SELECT	Allowed	ranger-acl	127.0.0.1	d1	staging
—		03/17/2019 06:32:54 PM	etluser	cl1_hive hive	weblog_staging/weblog/dob...@column	SELECT	Denied	ranger-acl	127.0.0.1	d1	staging
—		03/17/2019 06:32:33 PM	mark_bizdev	cl1_hive hive	claims_landing/claims_filed/...@column	SELECT	Denied	ranger-acl	127.0.0.1	d1	landing
—		03/17/2019 06:32:28 PM	admin	cl1_hive hive	claims_landing/claims_filed/...@column	SELECT	Denied	ranger-acl	127.0.0.1	d1	landing
95	3	03/17/2019 06:32:22 PM	etluser	cl1_hive hive	claims_landing/claims_filed/...@column	SELECT	Allowed	ranger-acl	127.0.0.1	d1	landing

- Shows zone of the accessed resource
- Audits can be filtered by zone
- Only policies in zone of the accessed resource are used to authorize

ROLE BASED ACCESS CONTROL

APACHE RANGER: ROLE BASED ACCESS CONTROL - INTRODUCTION

- Ranger policy model extended to support roles
- RBAC is widely used in enterprise applications & cloud environments
- Roles can be used in
 - resource-based authorization policies
 - tag-based authorization policies
 - data-masking policies
 - row-filtering policies
- Role management REST API

APACHE RANGER: ROLE BASED ACCESS CONTROL – ROLE ADMIN

Ranger

Access Manager

Audit

Security Zone

Settings

admin

Users/Groups/Roles

Users

Groups

Roles

Role List

Search for your roles

Add New Role

	Role Name	Users	Groups
	hr-admin	kate_hr	--
	admin	admin	--
	analyst	--	analyst
	hr-analyst	joe_analyst	--

APACHE RANGER: ROLE BASED ACCESS CONTROL - POLICY

Ranger

Access Manager

Audit

Security Zone

Settings

admin

Service Manager

cl1_hive Policies

Edit Policy

Policy Type

Masking

Add Validity Period

Policy ID

16

Policy Name *

mask.ccn show first 4

enabled

normal

Policy Label

Policy Label

Hive Database *

hortoniabank

Hive Table *

us_customers

Hive Column *

ccn

Description

Audit Logging

YES

Policy Conditions

No Conditions

Add

Mask Conditions :

Select Role

Select Group

Select User

Policy Conditions

Access Types

Select Masking Option

analyst

Select Groups

Select Users

Add Conditions

select

Partial mask: show first 4

CONDITIONS AT POLICY SCOPE

APACHE RANGER: CONDITIONS AT POLICY SCOPE - INTRODUCTION

- Conditions can now be set at policy scope, in addition to policy-item scope
- Simplifies use of conditions in policies
- Example use cases:
 - Policies specific to access cluster i.e. on-prem, cloud
 - Multiple policies for a given tag, for different tag-attribute values
i.e. PII type=email, PII: type=ccn

APACHE RANGER: CONDITIONS AT POLICY SCOPE - SAMPLE

Ranger

Access Manager

Audit

Security Zone

Settings

admin

Service Manager

cl1_tag Policies

Access

Masking

List of Policies : cl1_tag

Search for your policy...

Add New Policy

Policy ID	Policy Name	Policy Labels	Status	Audit Logging	Groups	Users	Action
1	EXPIRES_ON	—	Enabled	Enabled	public		
24	PII - access from cloud		Enabled	Enabled	public		
25	PII - email		Enabled	Enabled	csr		
26	PII - ccn		Enabled	Enabled	finance		

Ranger

Access Manager

Audit

Security Zone

Settings

admin

Service Manager

cl1_tag Policies

Edit Policy

Edit Policy

Policy Details :

Policy Type

Access

Policy ID

24

Policy Name *

PII - access from cloud

enablednormal

Policy Label

Policy Label

TAG *

PII

Description

Restrict access to PII data from cloud clusters.

Audit Logging

YES

Policy Conditions

Access cluster type? cloud

Add Validity Period

Access cluster type: cloud

cloudera

Privacera

© Cloudera, Inc. All rights reserved.

34

THANK YOU