4 ESSENTIAL STEPS FOR MANAGING SENSITIVE DATA

HORTONWORKS®

Privacera

SPEAKERS

Balaji Ganesan

CEO, Privacera

Srikanth Venkat

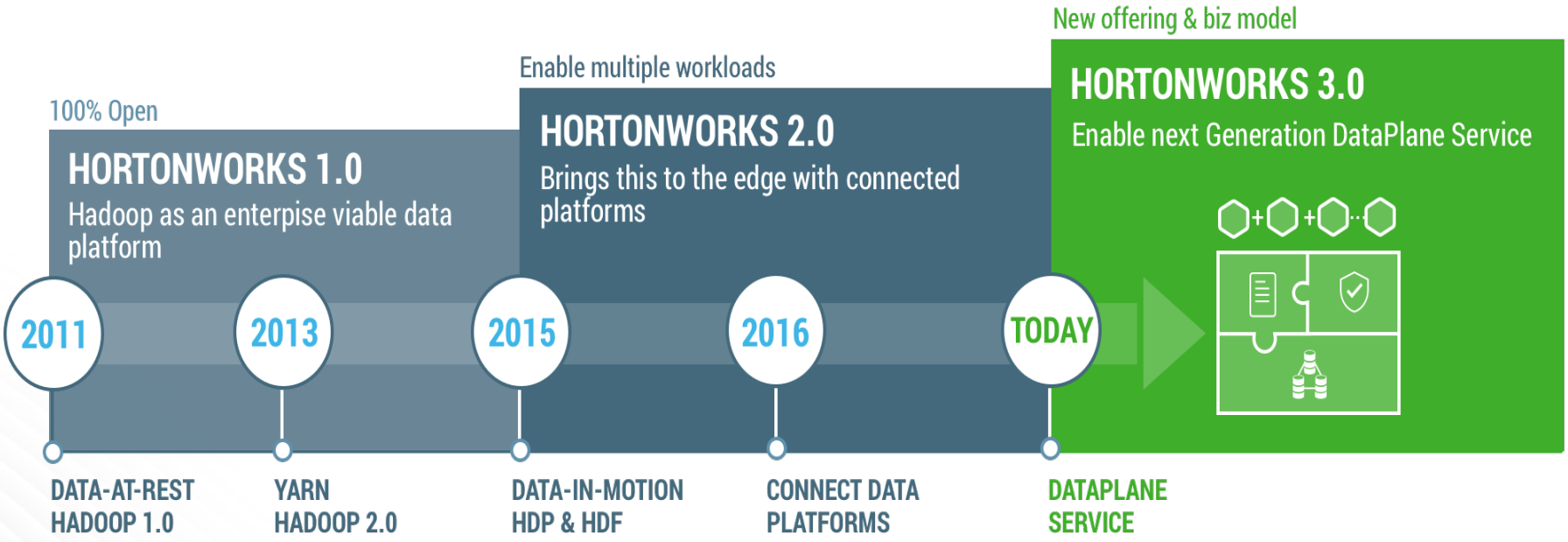
Senior Director, Product Management, Hortonworks

AGENDA

- Hortonworks Introduction
  - Security & Governance with Hortonworks
  - Sensitive Data Management Challenges
  - Hortonworks DataPlane Service
  - Demo (Data Steward Studio)
- Privacera Introduction
  - 4 steps in managing sensitive data
  - Representative scenarios & solutions
  - Demo (Privacera)
- Wrap up

About Hortonworks:

Enabling the Modern Data Architecture through consistent and continuous innovation



HORTONWORKS

Apache Ranger

Authorization	Ranger KMS	Auditing

• Centralized platform to define, administer and manage security policies consistently across Hadoop components
  • HDFS, Hive, HBase, YARN, Kafka, Solr, Storm, Knox, NiFi, Atlas
• Extensible Architecture
  • Custom policy conditions, user context enrichers
  • Easy to add new component types for authorization
	• Store and manage encryption keys
	• Support HDFS Transparent Data Encryption
	• Integration with HSM
	  • Safenet LUNA
		• Central audit location for all access requests
		• Support multiple destination sources (HDFS, Solr, etc.)
		• Real-time visual query interface

HORTONWORKS

Dynamic Row Filtering & Column Masking: Apache Ranger with Apache Hive

User 1: Joe
Location : US
Group: Analyst

Users from US Analyst group see data for US persons with CC and National ID (SSN) as masked values and MRN is nullified

EU HR Policy Admins can see unmasked but are restricted by row filtering policies to see data for EU persons only

User 2: Ivanna
Location : EU
Group: HR

Country	National ID	CC No	MRN	Name
US	xxxxx3233	4539 xxxx xxxx xxxx	null	John Doe
US	xxxxx7465	5391 xxxx xxxx xxxx	null	Jane Doe

Country	National ID	Name	MRN
Germany	T22000129	Ernie Schwarz	876452830A

Ranger Policy Enforcement

Query Rewritten based on Dynamic Ranger Policies: Filter rows by region & apply relevant column masking

Original Query:

SELECT country, nationalid, ccnumber, mrn, name FROM ww_customers

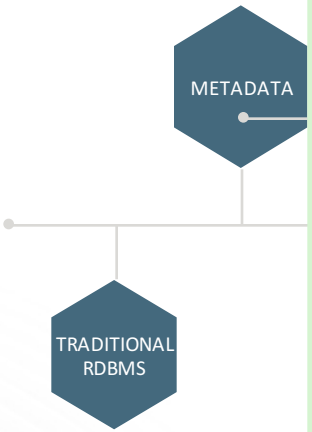
Original Query:

SELECT country, nationalid, name, mrn FROM ww_customers

Country	National ID	CC No	DOB	MRN	Name	Policy ID
US	232323233	4539067047629850	9/12/1969	8233054331	John Doe	nj23j424
US	333287465	5391304868205600	8/13/1979	3736885376	Jane Doe	cadsd984
Germany	T22000129	4532786256545550	3/5/1963	876452830A	Ernie Schwarz	KK-2345909

HORTONWORKS

Apache Atlas : Open Metadata & Governance



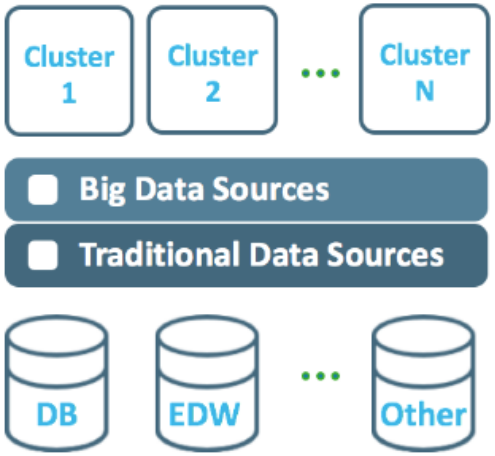
Vision:

Metadata-driven foundational governance services for enterprise data ecosystem

- Open frameworks and APIs
- Agile and secure collaboration around data and advanced analytics
- Reduce operational costs while extracting economic value of data

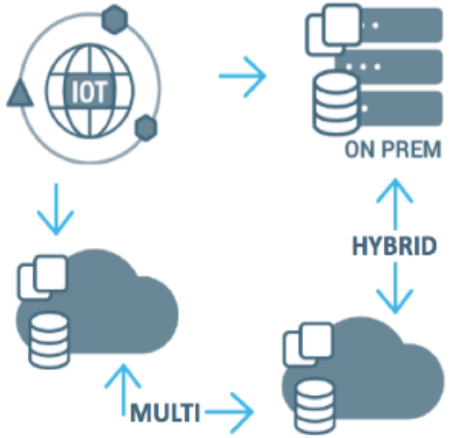
HORTONWORKS

Next Generation Data Problems



Data Is Spread Across Multiple Clusters and Data Sources

Store & Analyze Data From ERP/CRM, Systems, IoT/ Mobile Devices, Social Media, Geo Location etc.



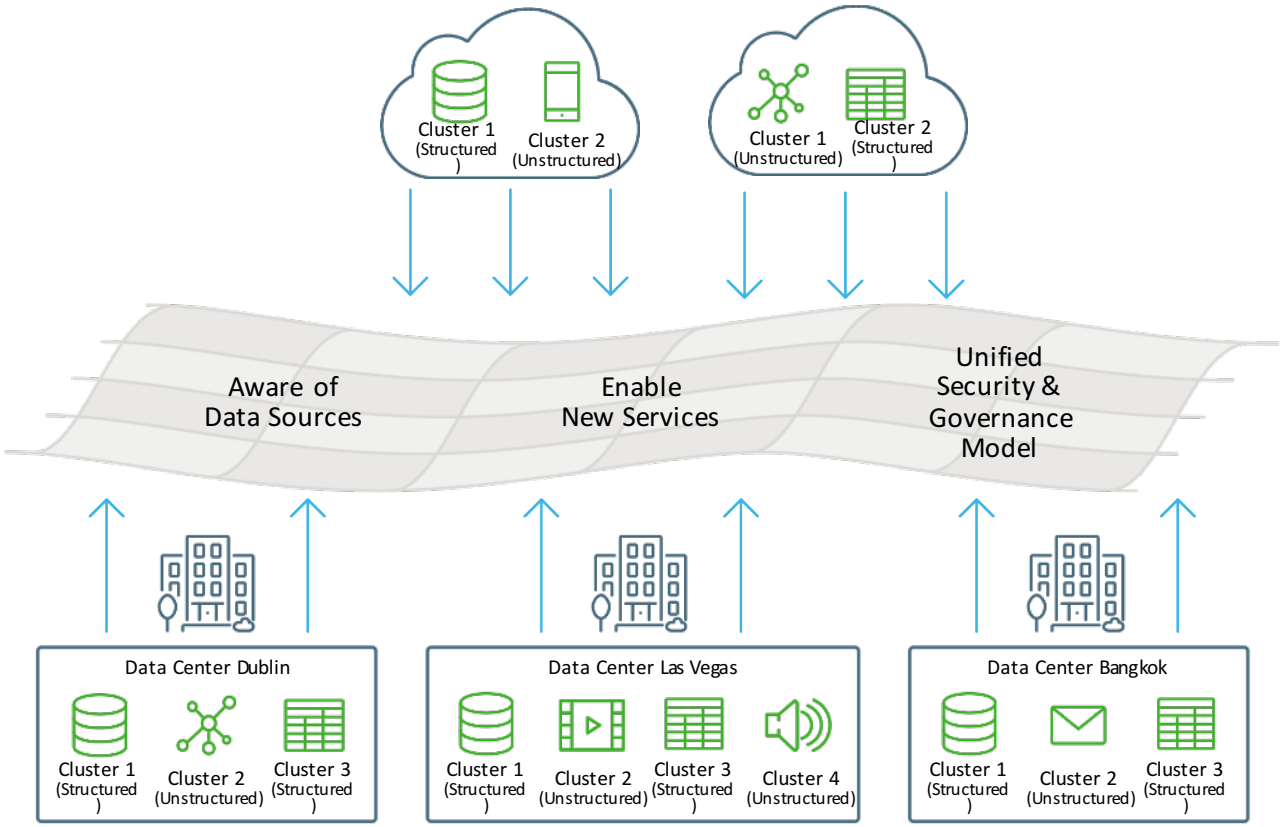
Some data is on-premise, rest in the cloud.

Moving data from cloud to on-premise & vice versa

Moving data between different clouds

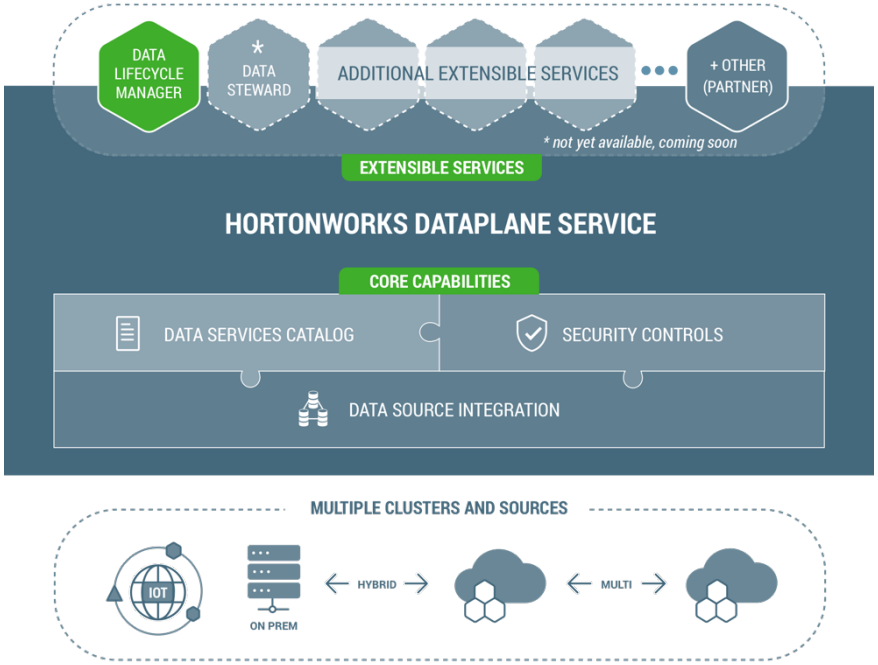
HORTONWORKS

What If…



HORTONWORKS

What is Hortonworks DataPlane Service?



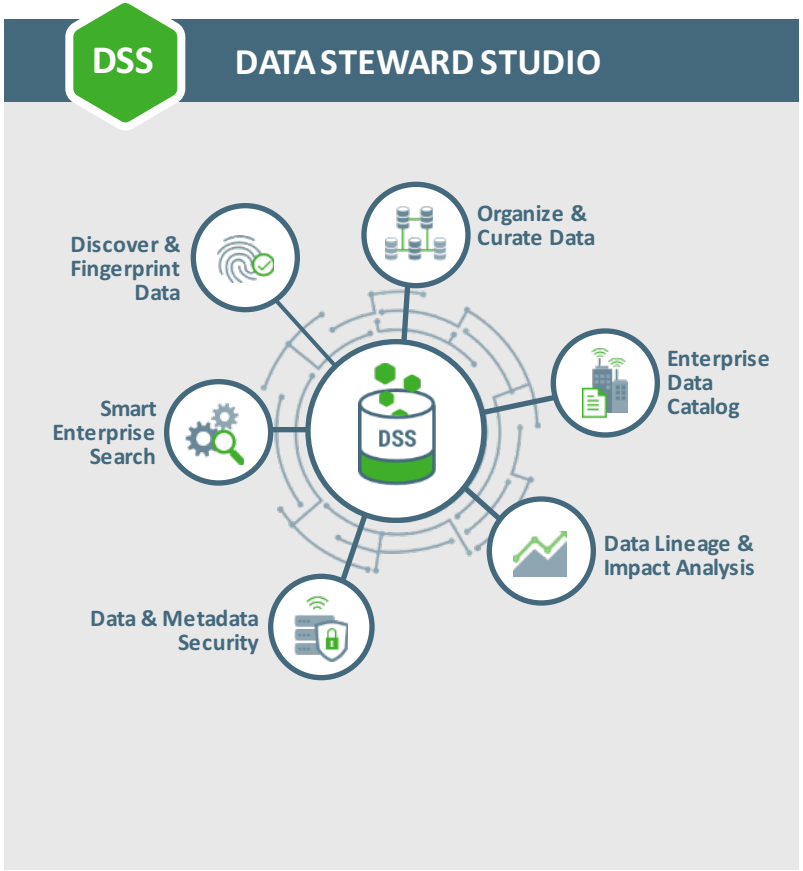
Hortonworks DataPlane Service

a platform with extensible data management services for:

- Addressing compliance and regulatory requirements for enterprise
- Providing consistent security & governance across data landscape
- Enabling centralized management of data assets
- Responsible data sharing and collaboration

HORTONWORKS

Hortonworks DataPlane Service: Extensible Services



Data Steward Studio (DSS)

Suite of capabilities that allows users to understand, secure, and govern data across enterprise data lakes

Ensure consistent security and governance for data assets **across tiers**

- Curate, discover and organize data assets
  based on business classifications, purpose, protections, relevance, etc.
- Govern proper usage and lineage of data assets
  to identify schema, classification and view lineage/data supply chain
- Understand and audit data asset security and use
  for anomaly detection, forensic audit/compliance & proper control mechanisms

…all across multiple types and tiers of data

Technical Preview Available

HORTONWORKS

Data Steward Studio (DSS)

CONSUMABILITY: Understand shape of Hive column data with statistical profiler, example: Profile shows box plot and histogram for distribution of column values

Data Steward / Asset Details

Datalake
cl1

Database
hortoniabank

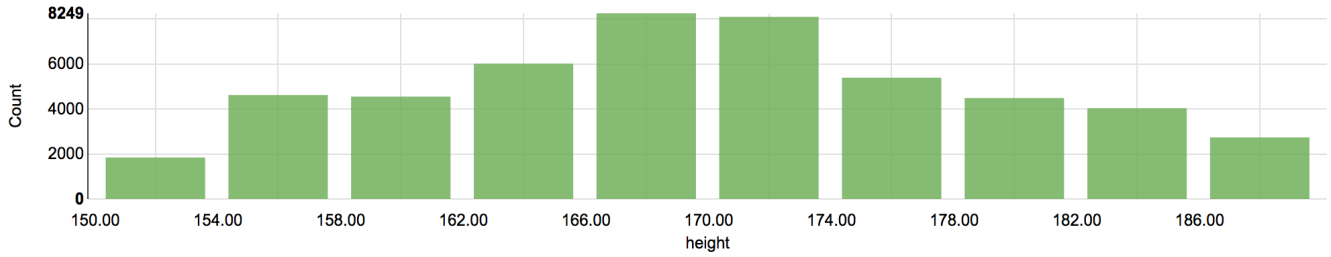
DETAILS LINEAGE POLICY AUDIT

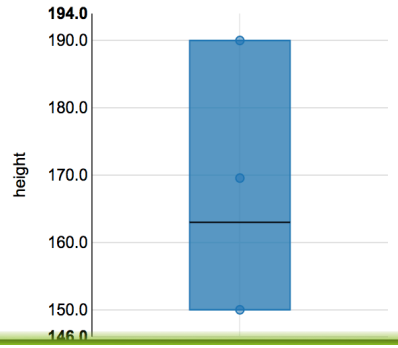
PROPERTIES TAGS SCHEMA

Next Profiler Schedule in : 41 minutes

	Name	Type	Unique Values *	Null Values	Max	Min	Mean	Comment
	streetaddress	string	49790	0				
	middleinitial	string	26	0				
	emailaddress	string	49970	0				
	age	int	67	0	85	19	52.17764	
	height	int	41	0	190	150	169.56316	

Profiled : 100% rows, 19 minutes ago





Data Steward Studio (DSS)

CONSUMABILITY: Data lineage shows complete chain of custody and downstream dependencies for an asset!

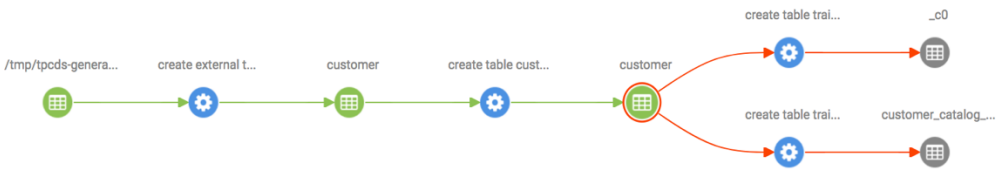
Data Steward / Asset Collection / Asset Details

CUSTOMER

HIVE

Datalake	Database	# of Rows
prod_east	tpcds_bin_partitioned_orc_2	144000

DETAILS LINEAGE POLICY AUDIT

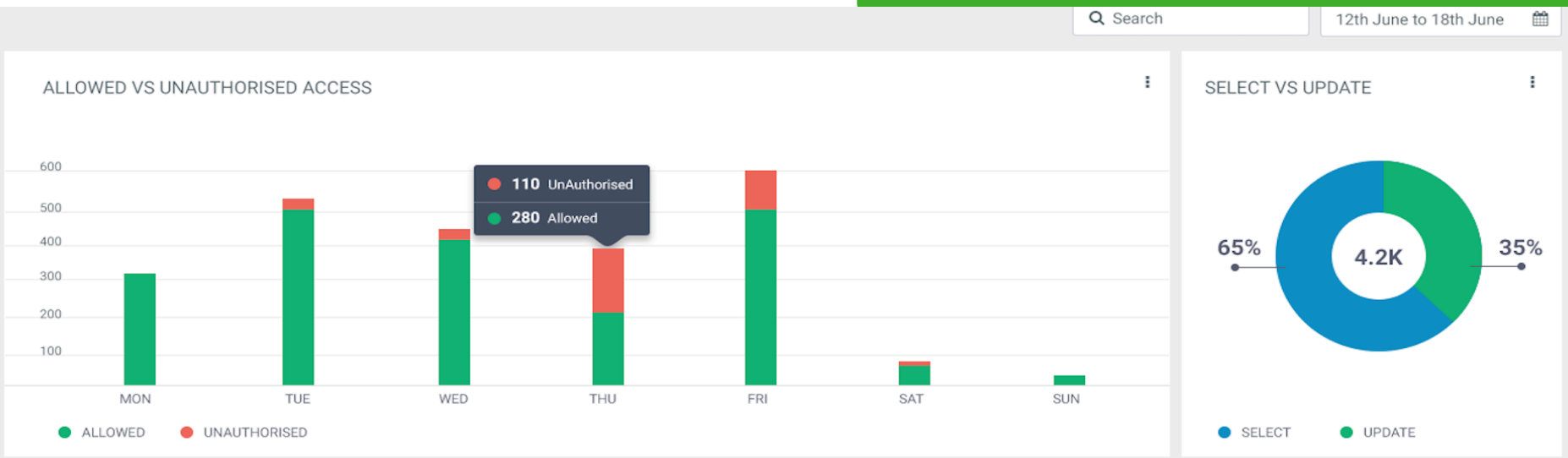


customer

owner:	centos
temporary:	false
lastAccessTime:	29 Sep 2017 06:54:35 PM
aliases:	
qualifiedName:	tpcds_text_2.customer...
description:	
viewExpandedTex	
tableType:	EXTERNAL_TABLE
createTime:	29 Sep 2017 06:54:35 PM
name:	customer
comment:	
partitionKeys:	
profileData:	typeName: hive_table_pr...
db:	tpcds_text_2

Data Steward Studio (DSS)

CONSUMABILITY: Audit Profiler shows both summarized views & patterns of access for a data asset.



Policy ID	Event Time	User	Access Type	Result	Access Enforcer	Client IP
19	07/28/2017 07:45:09 GMT	admin	UPDATE	ALLOWED	ranger-acl	172.27.25.135
19	07/28/2017 07:42:42 GMT	admin	UPDATE	ALLOWED	ranger-acl	172.27.25.135
19	07/28/2017 06:31:03 GMT	admin	CREATE	ALLOWED	ranger-acl	172.27.25.135

WEBINAR

PRIVACERA INTRODUCTION

ABOUT PRIVACERA

PLATFORM FOR DISCOVERING AND MANAGING SENSITIVE DATA

PARTNERS

HORTONWORKS
POWERING THE FUTURE OF DATA™

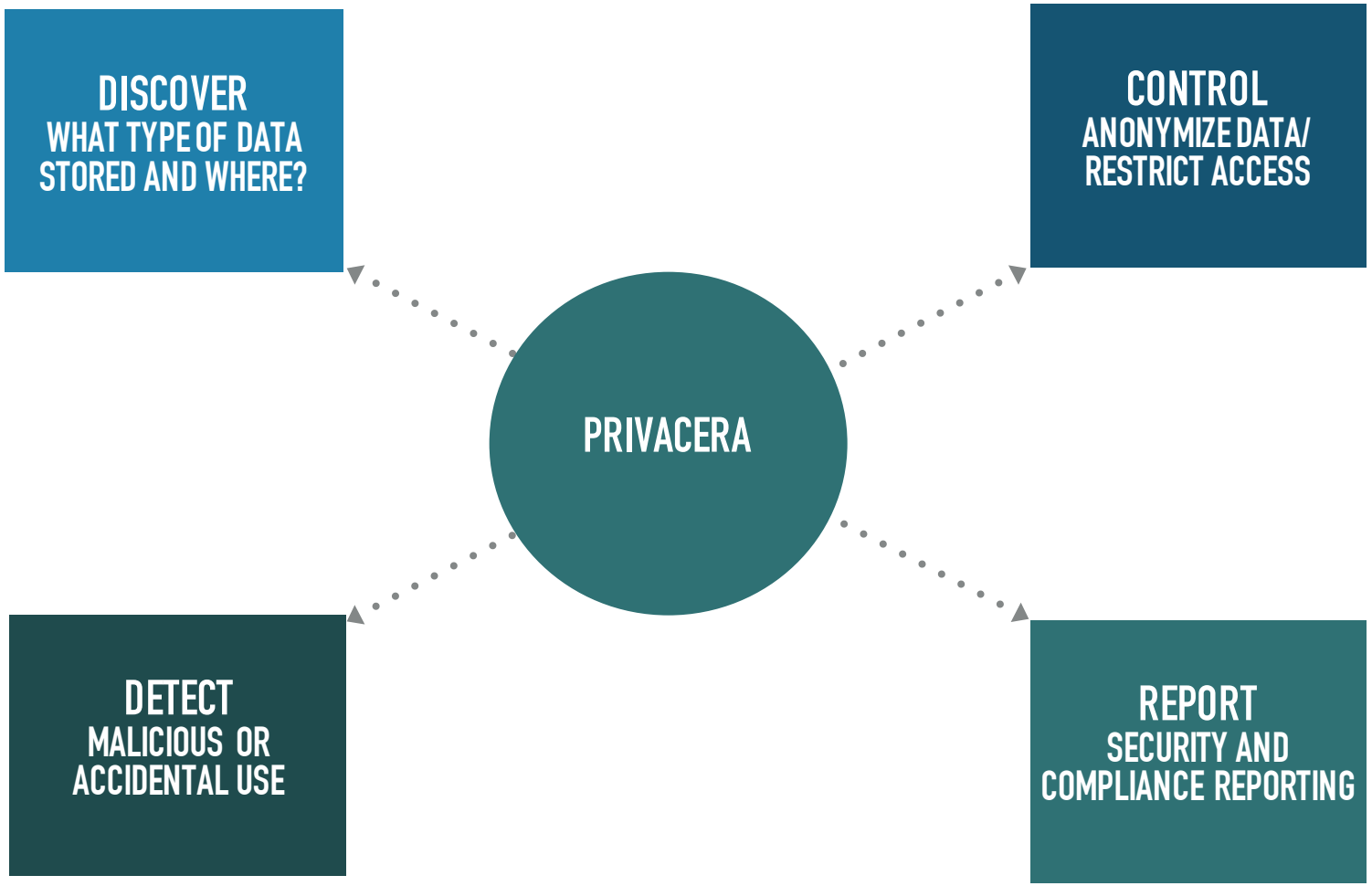
MAPR

cloudera

splunk>

GLOBAL

PLATFORM TO MANAGE SENSITIVE DATA



WEBINAR

STEPS TO MANAGE SENSITIVE DATA

4 STEPS FOR MANAGING SENSITIVE DATA

DATA DISCOVERY + ACCESS CONTROL + ANONYMIZATION + MONITORING

REPRESENTATIVE SCENARIO – FINANCIAL SERVICES

Multiple systems

Multiple formats

DATA LAKE

Sensitive data cannot be shared with users

INGESTION	STORAGE AND PROCESSING	DOWNSTREAM SYSTEMS

SOLUTION – PRIVACERA AUTOMATED DATA DISCOVERY

Discover and classify data during ingest or at rest

Standard rules combined with machine learning

Classification/ tags pushed to Atlas

INGESTION

STORAGE AND PROCESSING

STORAGE AND PROCESSING

REPRESENTATIVE SCENARIO – HEDGE FUND

Stock Info

Proprietary
Confidential data

DATA LAKE

Access to sensitive
data is restricted

Data Scientist

INGESTION	STORAGE AND PROCESSING	DOWNSTREAM SYSTEMS

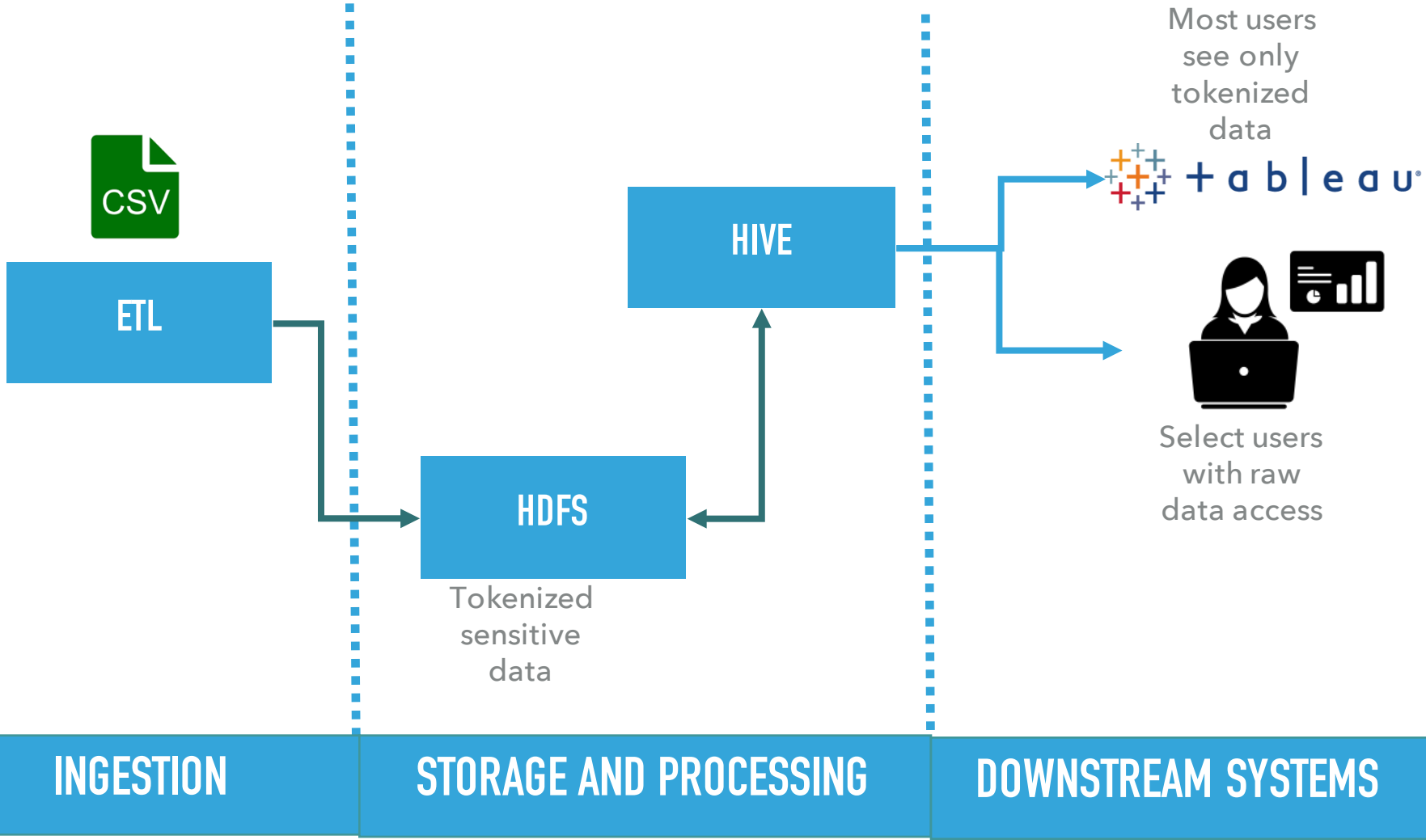
SOLUTION – TAG BASED ACCESS CONTROL

Simplify policies by managing at tag level

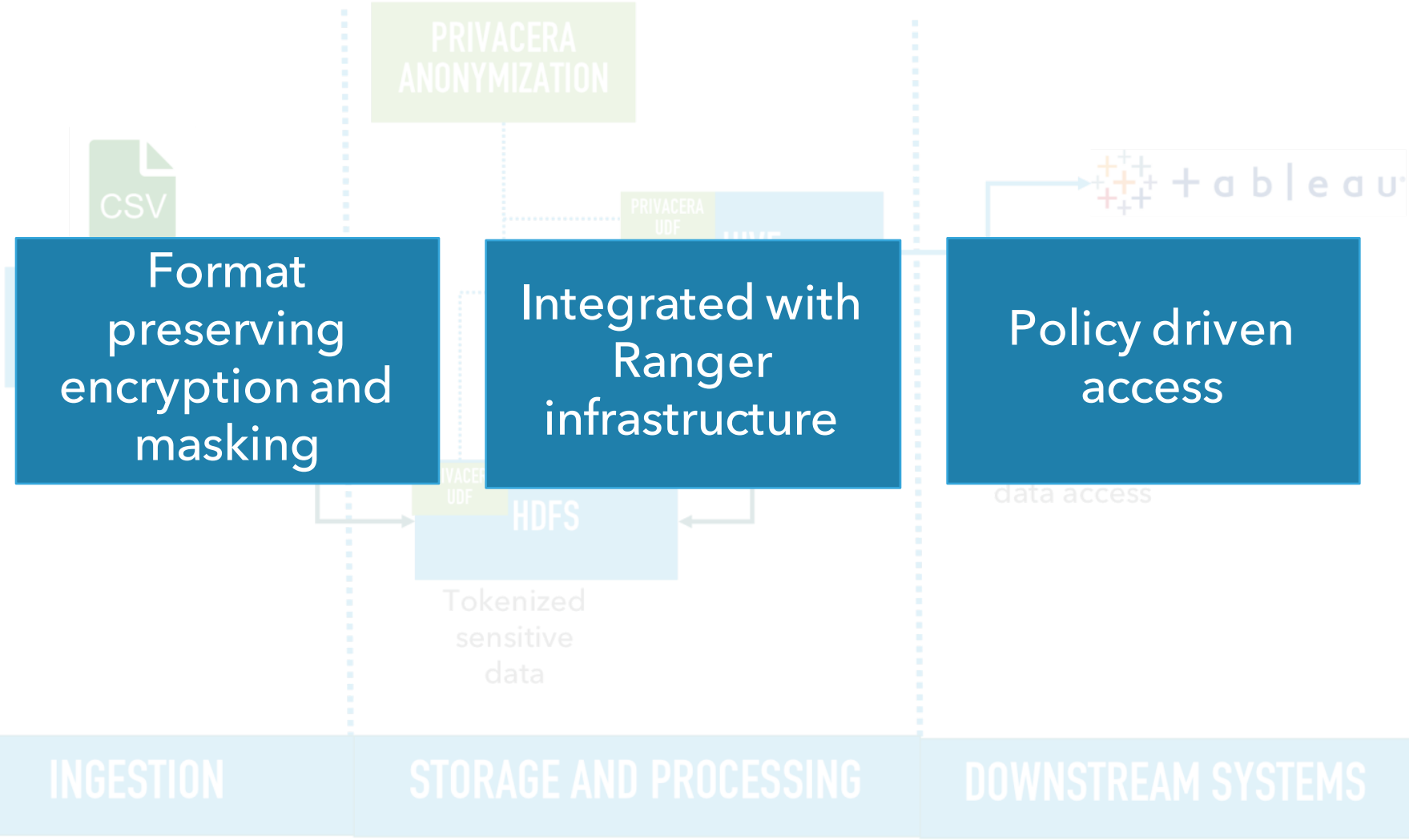
Tag attributes such as expiration date

Metadata updated by Privacera

REPRESENTATIVE SCENARIO – HEALTHCARE



SOLUTION – PRIVACERA ANONYMIZATION

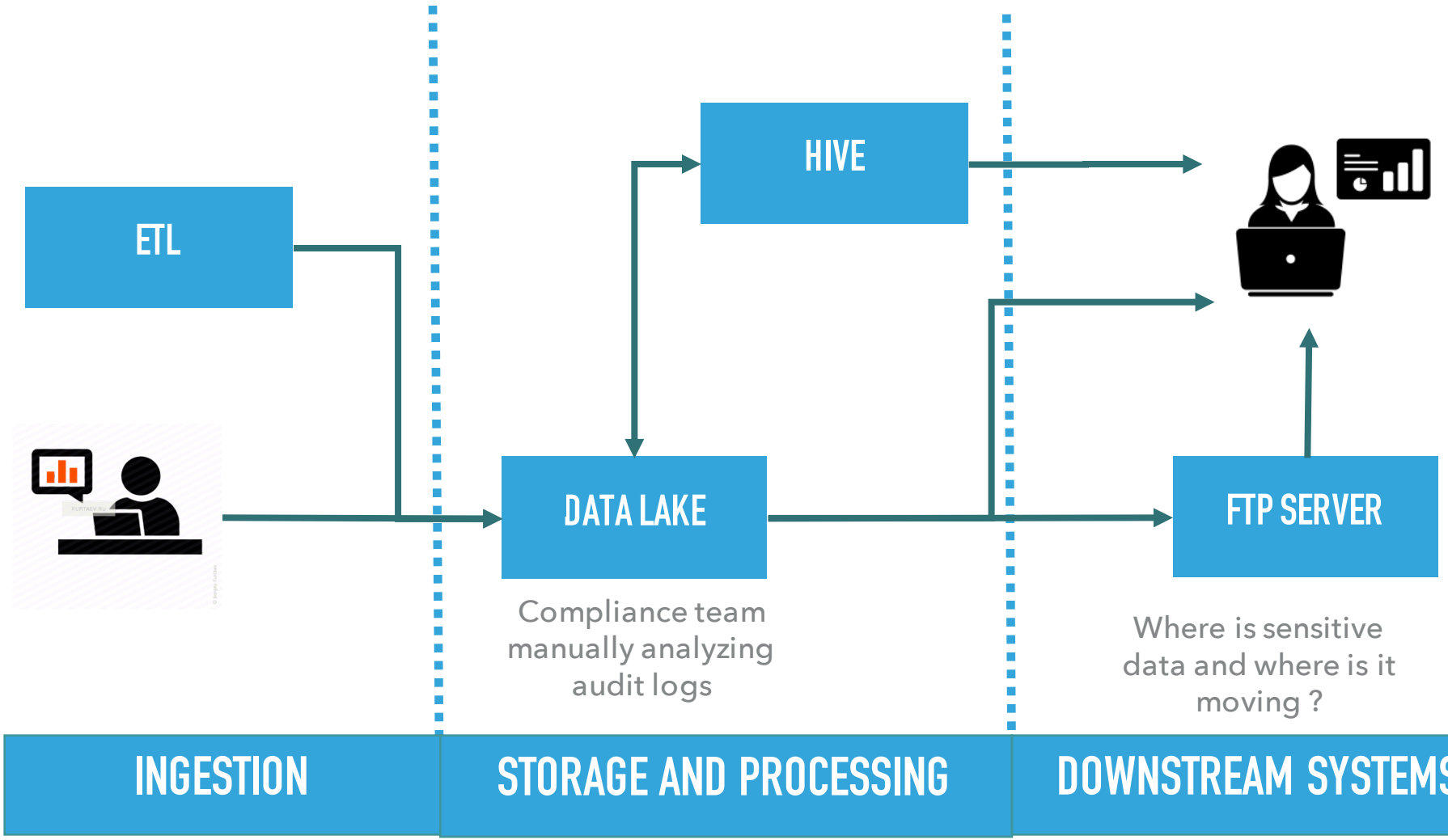


Format preserving encryption and masking

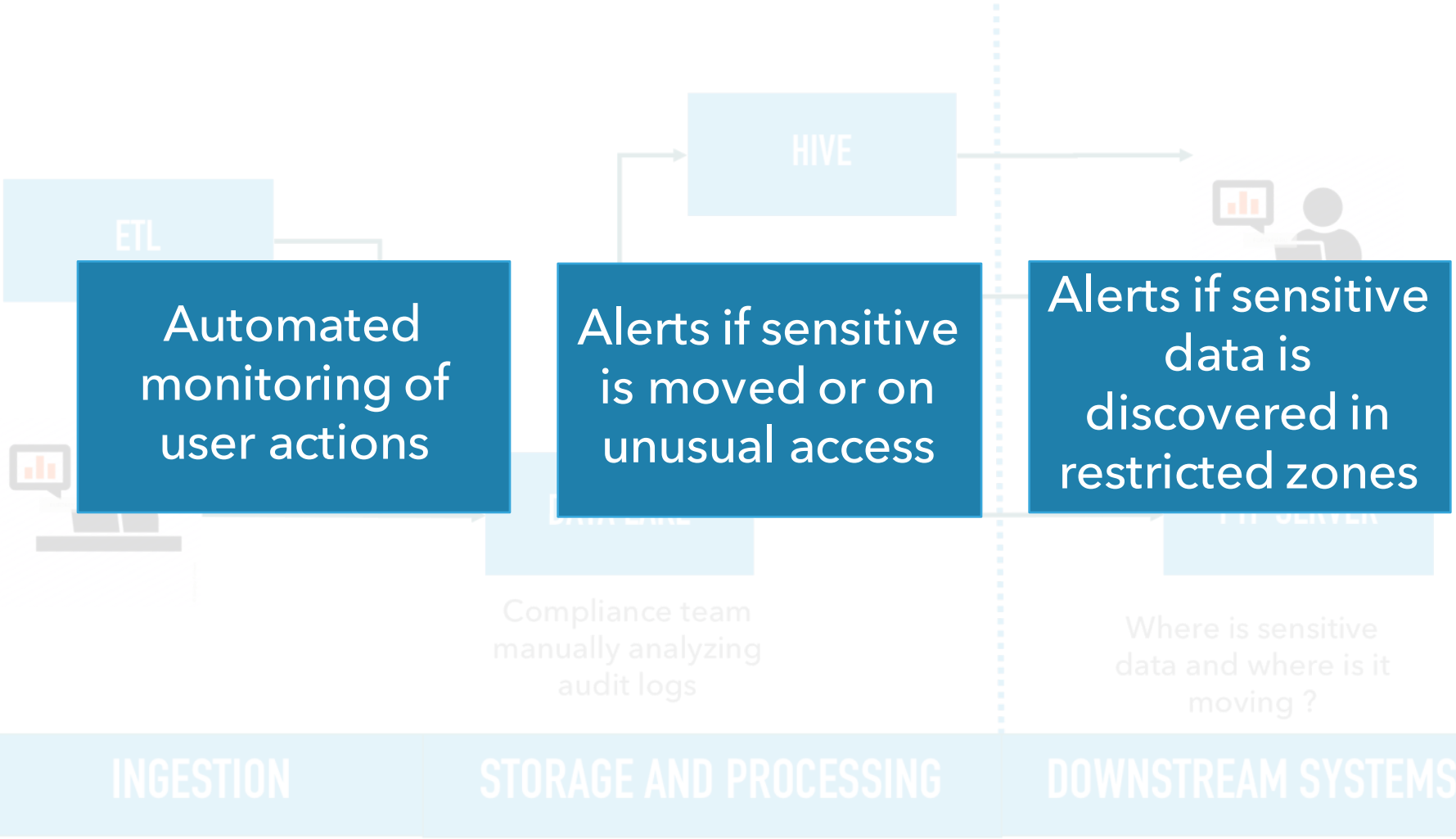
Integrated with Ranger infrastructure

Policy driven access

REPRESENTATIVE SCENARIO – FINANCIAL SERVICES



SOLUTION – PRIVACERA MONITORING



WEBINAR

DEMO

SUMMARY

- Understand your data before expanding your data lake
- Invest in automated classification and centralized metadata
- Manage access to user by data classification
- Anonymize data to reduce exposure
- Monitor the use of data, “trust but verify”.
- Data plane provides next generation for tools for hybrid data infrastructure

questions@privacera.com

QUESTIONS ?