Security Readiness for Deployment of Key Financial Applications in Hadoop

PRESENTERS:

Kenton Davis
Senior Director
BIAS Corporation

Ken Hall
Architect
SunTrust

BIAS

ORACLE Platinum Partner

SUNTRUST

About BIAS Corporation

Who We Are...

- **Founded in 2000**
- **Distinguished Oracle Leader**
  - Technology Momentum Award
  - Portal Blazer Award
  - Titan Award – Red Stack + HW Momentum Awards
  - Excellence in Innovation Award
- **Management Team is Ex-Oracle**
- **Location(s):** Headquartered in Atlanta; Regional office in Washington D.C.; Denver; Northern and Southern California; Charlotte, NC; Offshore – Hyderabad, Bangalore and Chennai, India
- **250 U.S. employees & contractors, 100 India employees, average with 10+ years of Oracle experience**
- **Inc.500|5000 Fastest Growing Private Company in the U.S. for the 7th Time**
- **Voted Best Place to work in Atlanta for 3rd year**
- **Top 10 Healthiest Workplace in Atlanta Business Chronicle**
- **33 Oracle Specializations spanning the entire stack**

About SunTrust Banks, Inc.

Our Presence*

Atlanta Headquarters	**25,324** Employees
$189.9 billion Total Assets	**$140.5 billion** Total Deposits
$151.7 billion Assets Under Advisement	**$141.8 billion** Mortgage Servicing Portfolio
4.2 million Consumer Households	**551,000** Business Clients
1,444 Branches	**2,163** ATMs

Primary Market Areas

Florida, Georgia, Maryland, North Carolina, South Carolina, Tennessee, Virginia, and the District of Columbia.

SUNTRUST

About the Speakers

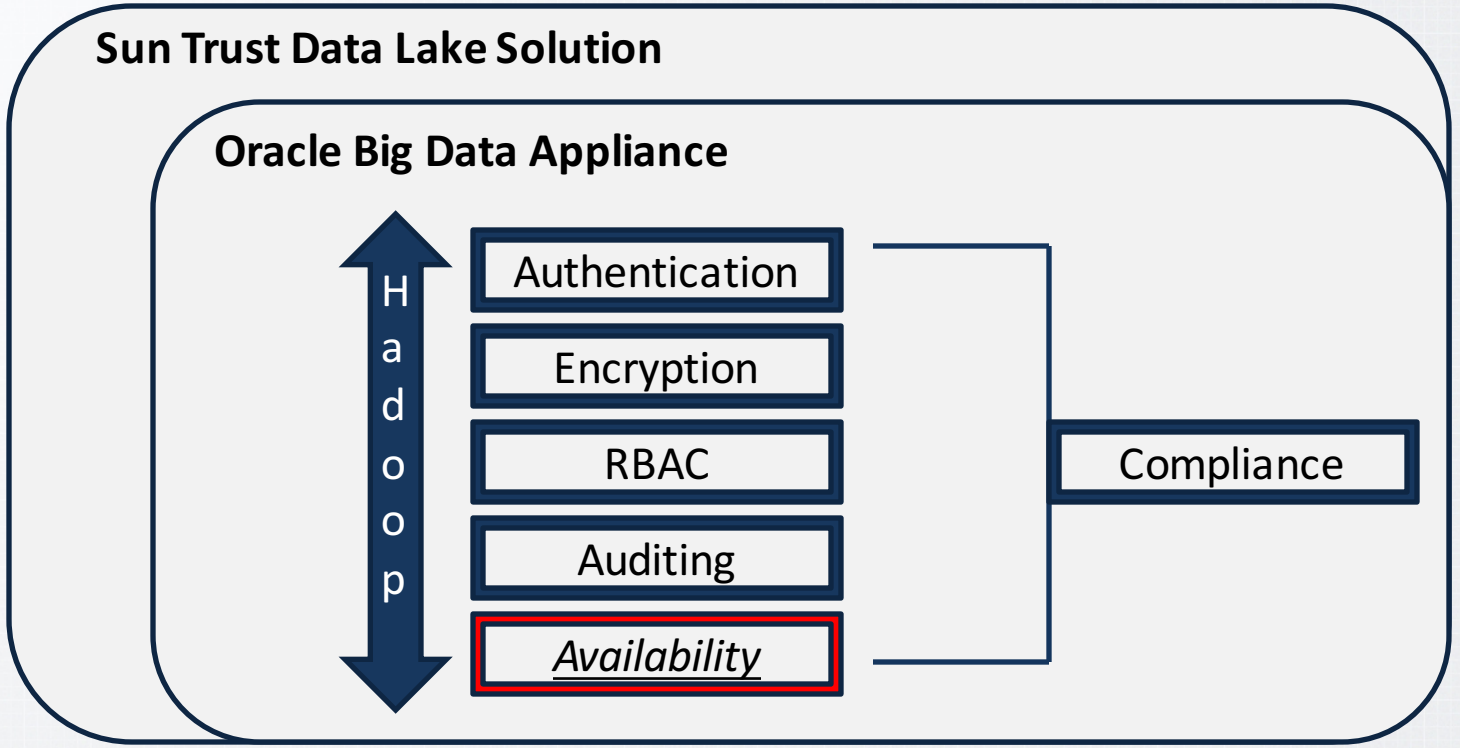
Ken Hall

Architect, SunTrust Banks, Inc.

Kenton Troy Davis

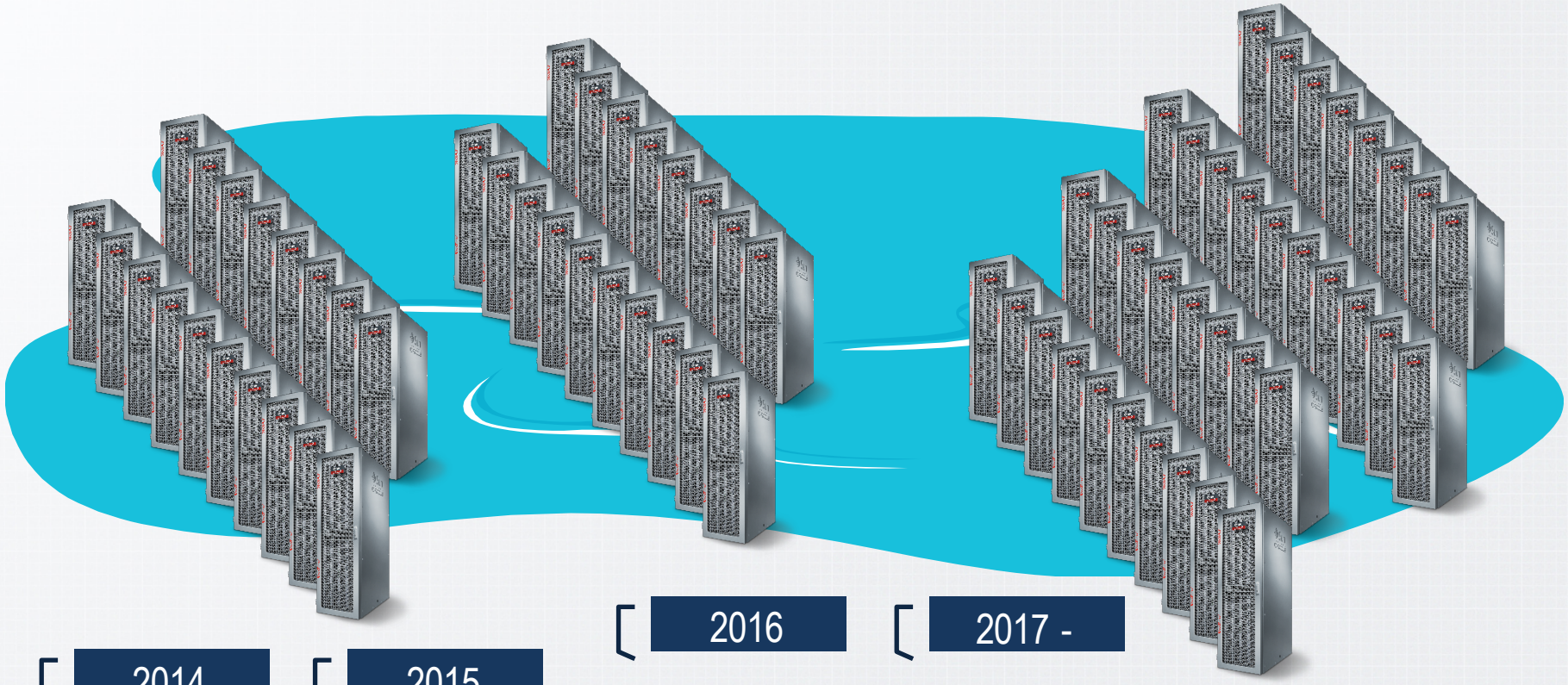
Senior Director & Enterprise Architect, BIAS Corporation

Agenda


Sun Trust Data Lake Solution
Oracle Big Data Appliance
Hadoop
Authentication
Encryption
RBAC
Auditing
Availability
Compliance

SunTrust Data Lake Solution

Projected to be Largest Oracle Big Data Appliance Implementation at a Bank



2014	2015	2016	2017 -
18 Node POC	66 Nodes	192 Nodes	2-3x growth currently projected

Business Goals of the Data Lake

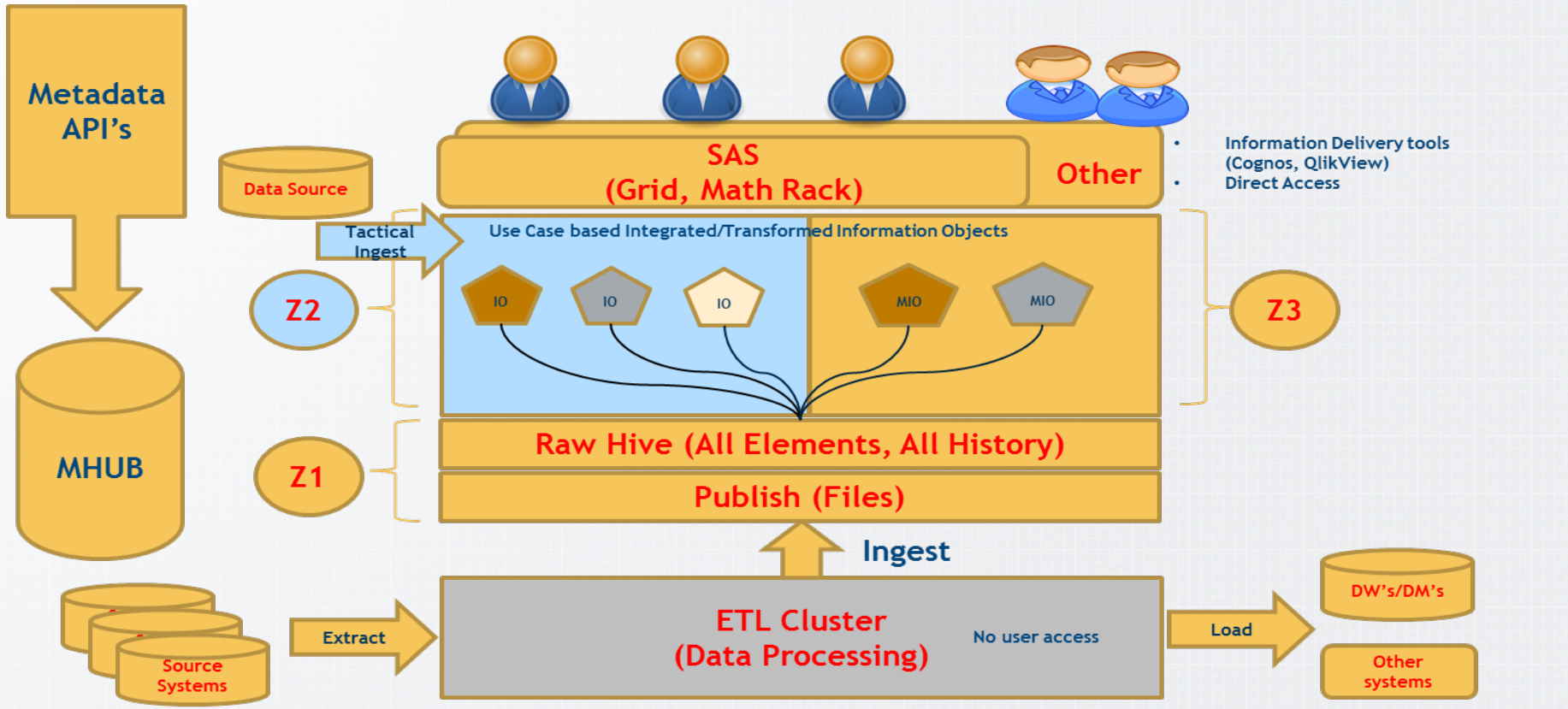
The challenges of increased regulatory pressure, cost efficiency, and speed to market with offerings based on client data insights drive our EDO Priorities.

Transforming our approach to enterprise data

Grow	Increase the quality, breadth and depth of data we store for our analytics and modeling communities.	Enable teammates to access volumes of information previously unreachable.	Derive insights from the volumes and quality of data now accessible to drive process, product, pricing and other innovations.
Reduce Cost	Remove redundancies and increase efficiency in how we store our data.	Reduce the number of times we move our data to clean it and stage it for use.	Remove complexity to enable users to focus on data analysis. Streamline and rationalize reports to reduce overhead and enhance BIO reporting.
Reduce Risk	Establish definitions and standards for the data we have so we know what we have, where it sits, and that it is high quality.	Provide a trusted and simpler way to access the data we have so it can be used more easily, with confidence.	Build sources of "**Truth**" to improve regulatory reporting and reduce fraud/ operational losses

Business Value

Data Lake Architecture

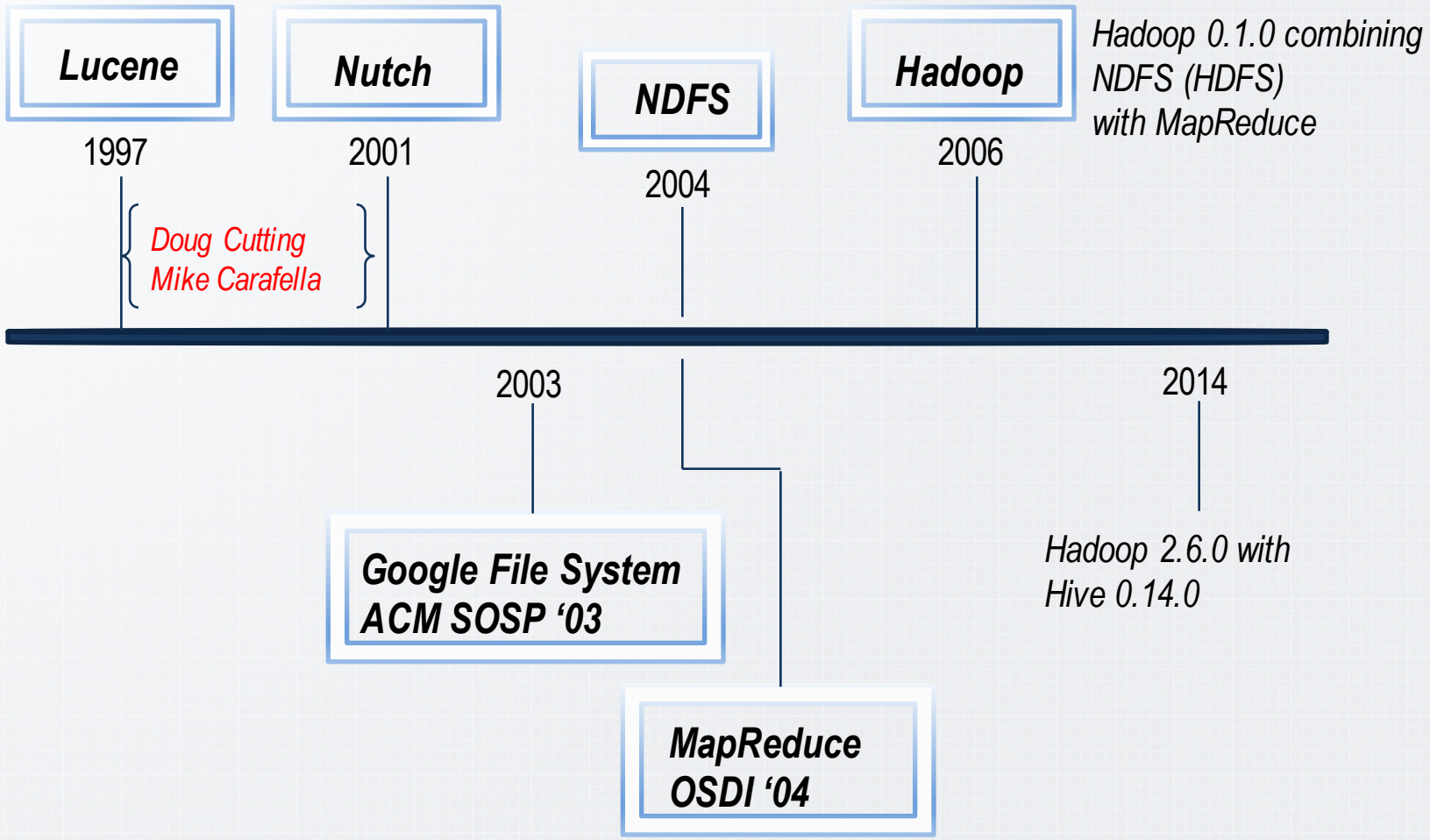


Growth Projections

Assumed consistent growth, Uncompressed estimates, Not including HDFS replication	2015	2016	2017	2018+
Social Media	23.00	23.00	23.00	23.00
IT Operational Data	11.50	11.50	11.50	11.50
Documentation, Images, Cheques Images (ECM)	57.50	57.50	57.50	57.50
Third Party Data Sources (700 Sources); Reference/ Bureau Quarterly	50.60	50.60	50.60	50.60
Bureau	8.05	8.05	8.05	8.05
Total Volume (TB)	**142.57**	**323.94**	**505.31**	**695.75**

Brief review of Hadoop and Hive

Early Timeline of Hadoop and Hive



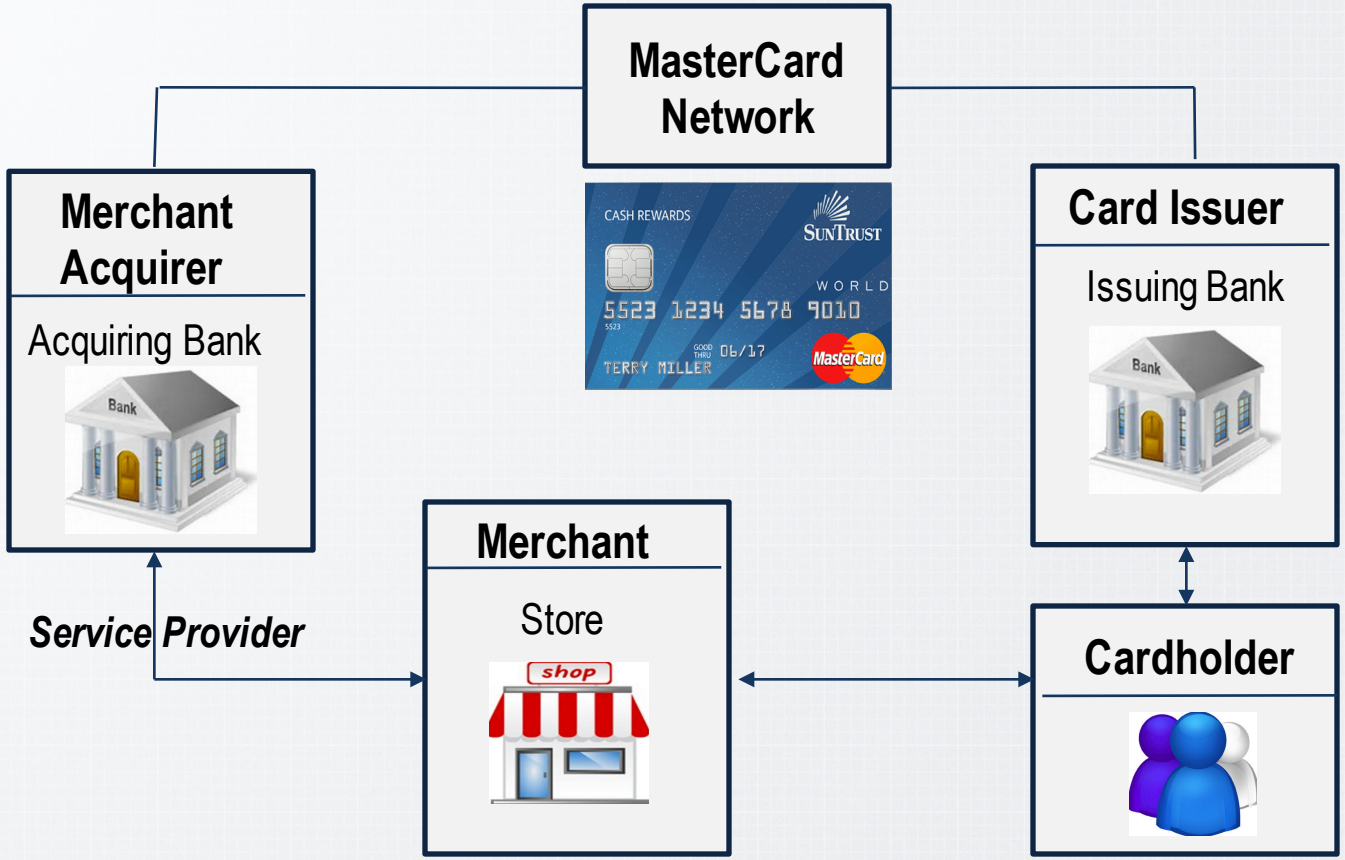
Some Apache Hive Features

- Offers the ability to query and summarize HDFS data using SQL-like (HiveQL) syntax
- Supports multiple execution engines within YARN (MR, Tez, and Spark)
- Enables the use of compression and indexing
- Supports different storage types (e.g. text, HBASE, Avro, Parquet, ORC)
- Incorporates metadata repositories (useful for data governance)
- Supports User-Defined Functions (useful for custom, security use cases)

Security: Regulation and Compliance

SOC2

Credit Card Transaction Parties



https://www.suntrust.com/personal-banking/credit-cards

Payment Card Industry Data Security Standard

- PCI-DSS is a financial industry mandate for protecting consumer credit card data.
- Acquiring banks are required to comply with PCI-DSS as well as validate compliance via audits.
- Issuing banks are not required to validate compliance via audits but must secure data in a PCI-DSS compliant manner.
- When a security breach occurs, all parties can be investigated and subject to penalties.

Payment Card Industry Data Security Standard

From PCI-DSS Data Security Standard V3.2 – April 2016:

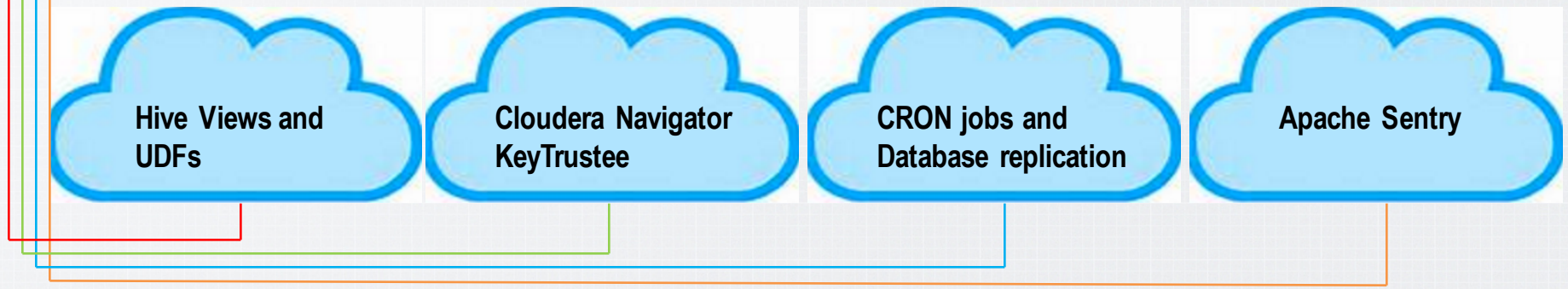
Goals	PCI DSS Requirements
Build and Maintain a Secure Network and Systems	1. Install and maintain a firewall configuration to protect cardholder data 2. Do not use vendor-supplied defaults for system passwords and other security parameters
Protect Cardholder Data	3. Protect stored cardholder data 4. Encrypt transmission of cardholder data across open, public networks
Maintain a Vulnerability Management Program	5. Protect all systems against malware and regularly update anti-virus software or programs 6. Develop and maintain secure systems and applications
Implement Strong Access Control Measures	7. Restrict access to cardholder data by business need to know 8. Identify and authenticate access to system components 9. Restrict physical access to cardholder data
Regularly Monitor and Test Networks	10. Track and monitor all access to network resources and cardholder data 11. Regularly test security systems and processes
Maintain an Information Security Policy	12. Maintain a policy that addresses information security for all personnel

Protect stored cardholder data

- Mask the Primary Account Number (PAN) such that at most only the first six digits and the last four digits are displayed.
- If a full unmasked PAN needs to be persisted, then it must be saved in encrypted form at rest.
- Documented procedures must exist for key management processes used for strong cryptography – e.g. for backup, key storage, key rotation *(section 3.6 sub controls)*, key access, etc.
- Principle of least privilege *(section 7)* applies by limiting data access according to which business groups 'need to know'.

Protect stored cardholder data

- Mask the Primary Account Number (PAN) such that at most only the first six digits and the last four digits are displayed.
- If a full unmasked PAN needs to be persisted, then it must be saved in encrypted form at rest.
- Documented procedures must exist for key management processes used for strong cryptography – e.g. for backup, key storage, key rotation *(section 3.6 subcontrols)*, key access, etc.
- Principle of least privilege *(section 7)* applies by limiting data access according to which business groups 'need to know'.



Kerberos Authentication

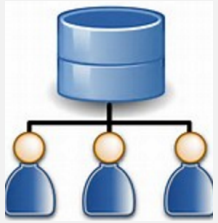
Kerberos Authentication

- Users (UPNs) and Services (SPNs) are uniquely identified as principals within an authentication domain called a realm.
- Users authenticate against the Key Distribution Center (KDC) to receive a ticket-granting ticket (TGT).
- Users submit the TGT to request access to a service.
- The KDC grants the user a service ticket having an expiration date and a maximum renewable period.

* From SAS Documentation in use at customer deployment

Kerberos Authentication

Active Directory KDC



hive/bdanode4.suntrust.com@CORP.SUNTRUST.COM

kenhall/etl-yarn-gateway.suntrust.com@CORP.SUNTRUST.COM

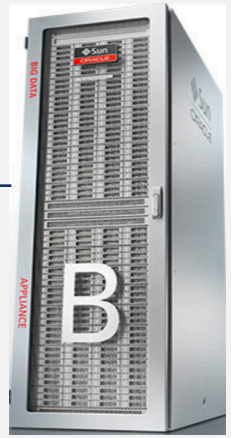
JDBC

User authenticates against KDC
Receives TGT
Kerberos cache populated

GSS APIs
Credentials
Delegation

HiveServer2

Node 4

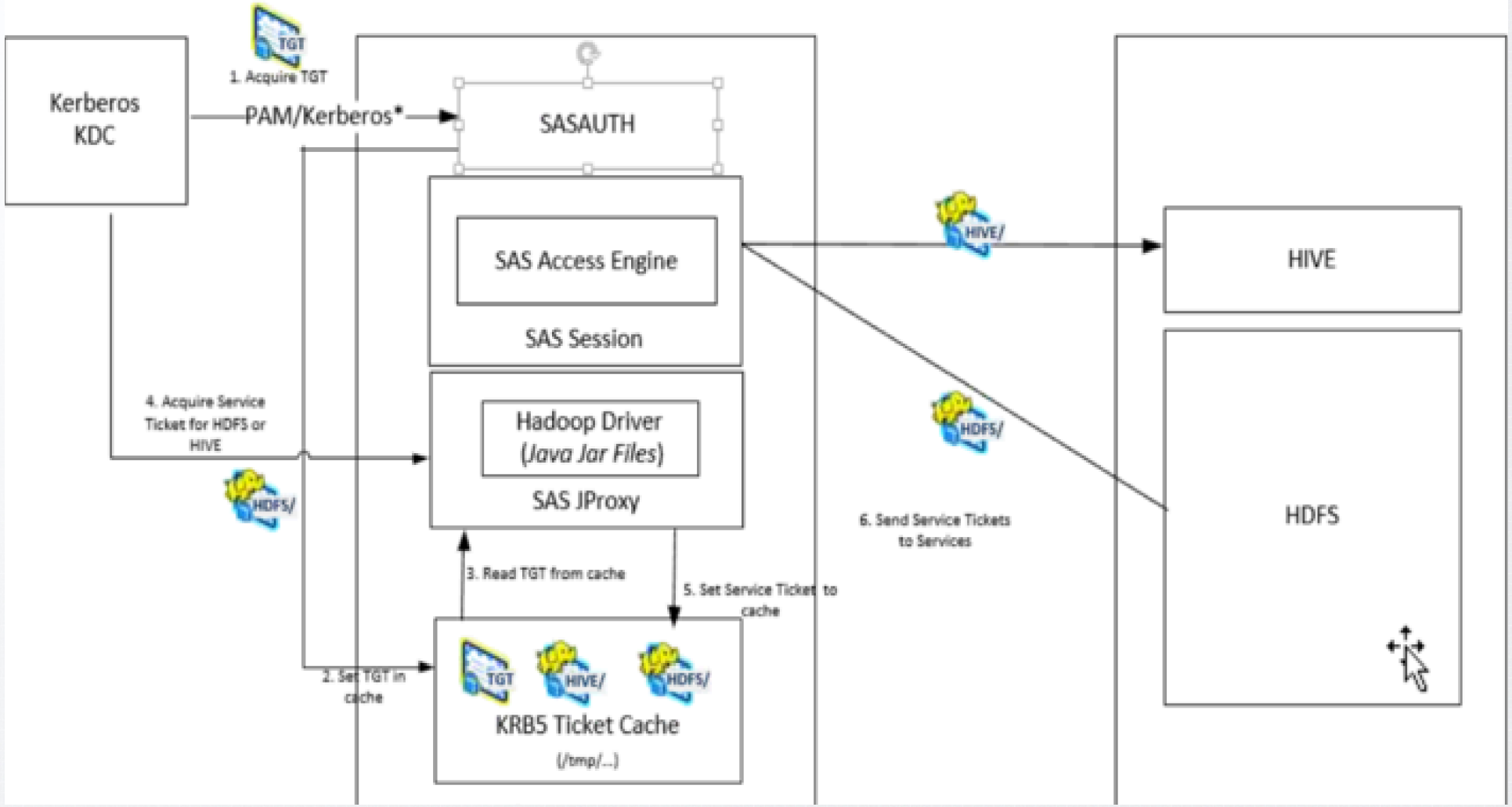


Cloudera Manager creates keytabs for Hadoop services

User requests service
Receives TGS to access Hive

CORP.SUNTRUST.COM Domain

SAS Access / Kerberos Example



* From SAS Documentation in use at customer deployment

Encryption

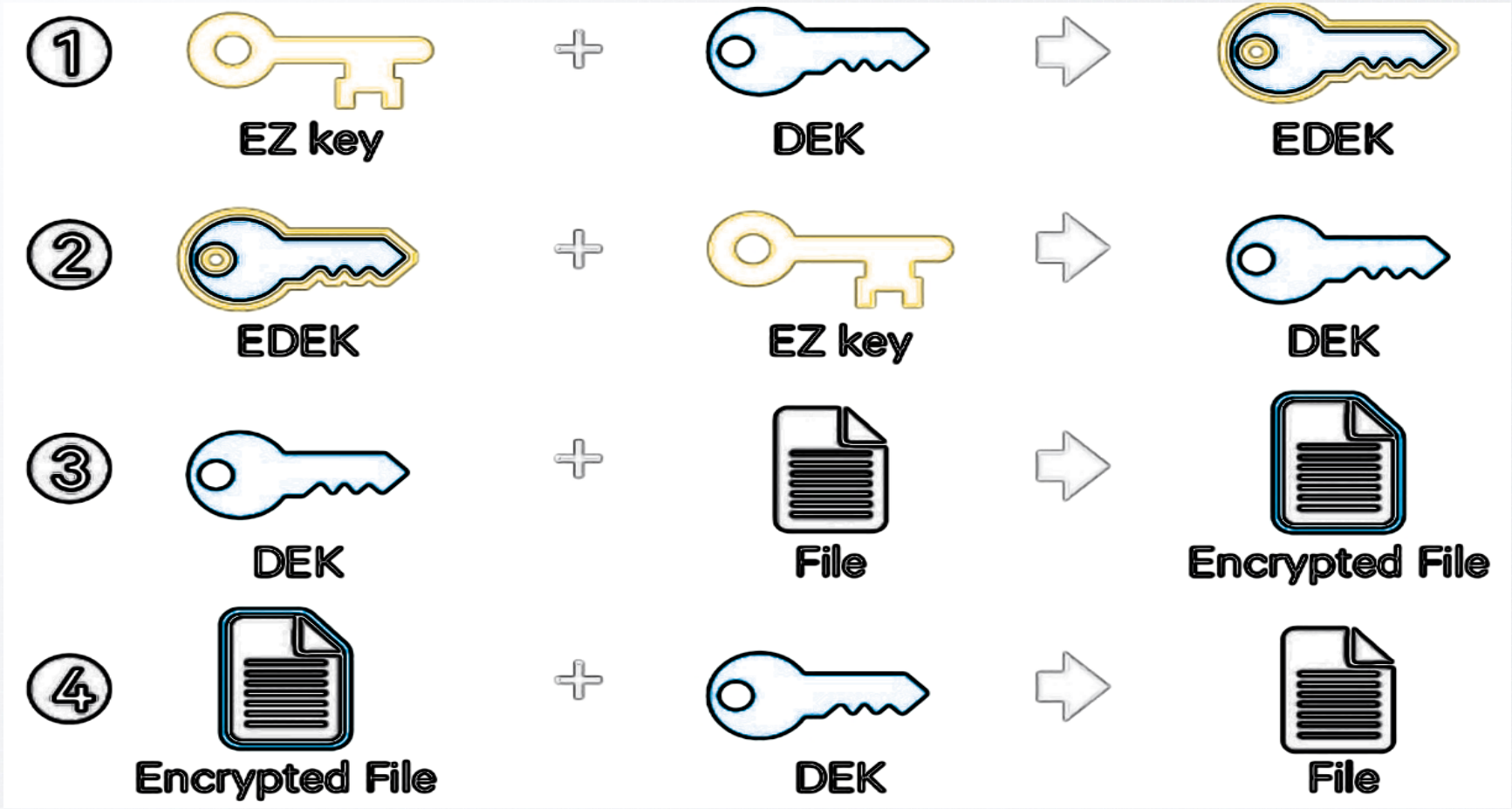
HDFS Transparent Data-at-Rest Encryption

- ELT/ETL processes at the Bank land files inside of an encryption zone within HDFS.
- An encryption zone is an HDFS directory in which every file and every subdirectory is required to contain encrypted content only.
- The encryption zone itself has a key called the EZ key.
- Each file in the encryption zone has its own key called a Data Encryption Key (DEK).

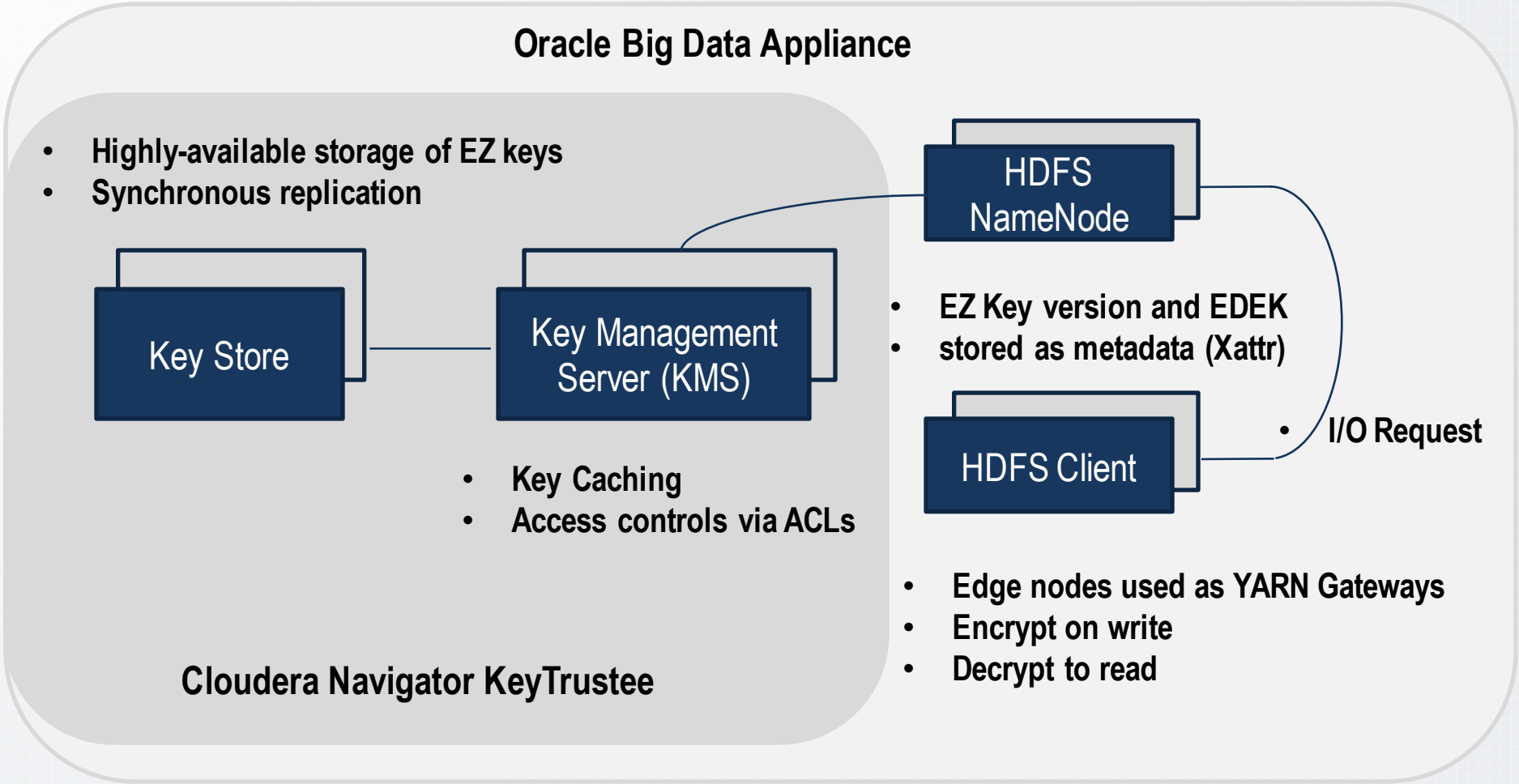
HDFS Transparent Data-at-Rest Encryption

- When applied to an HDFS file, the DEK encrypts the file enabling it to be securely stored.
- When applied to an encrypted HDFS file, the DEK decrypts the file enabling it to be read.

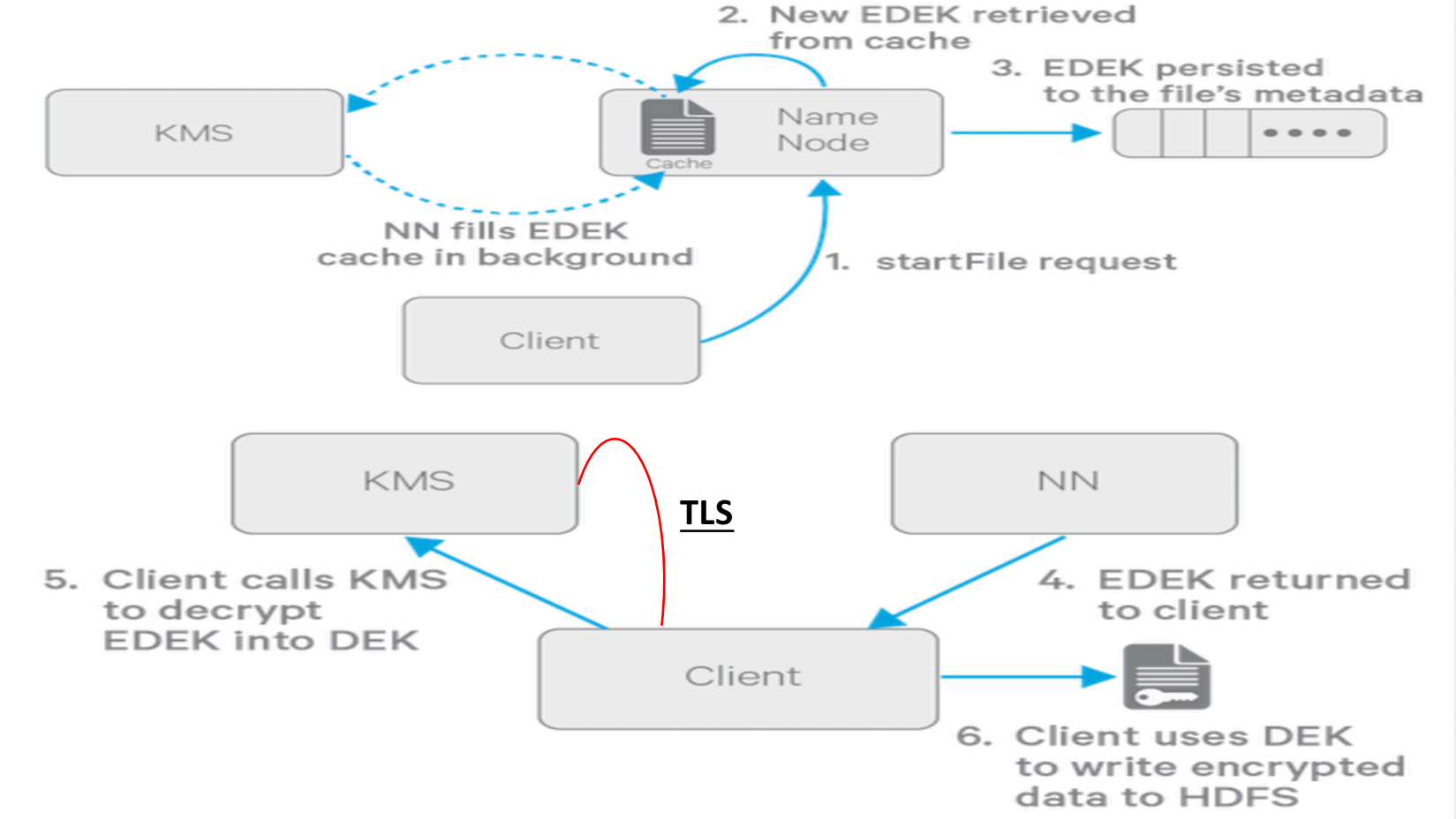
HDFS Transparent Data-at-Rest Encryption



HDFS Transparent Data-at-Rest Encryption



HDFS Transparent Data-at-Rest Encryption



* Modified based upon diagram from http://blog.cloudera.com/blog/2015/01/new-in-cdh-5-3-transparent-encryption-in-hdfs/

Key Management Subcontrols

Key Rotation Requirements

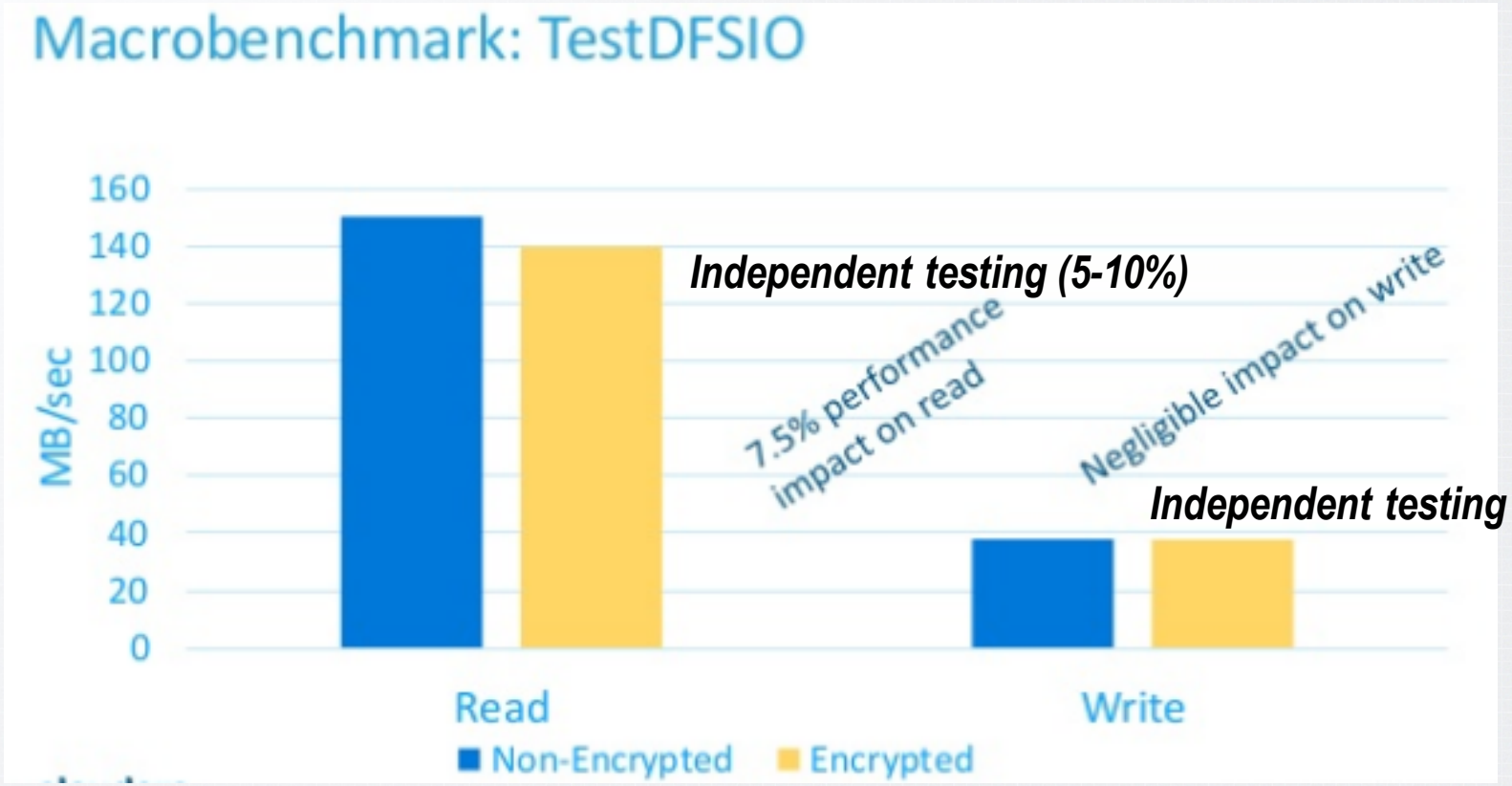
- An encryption key can have multiple versions
- Key rotation can be achieved by increasing the version.
- Per-file key rotation does not need to exhaustively re-encrypt all existing files.
- Instead, each file's existing DEK can simply be re-encrypted using the newer encryption zone key.

Key Management Subcontrols

Separation of Duties

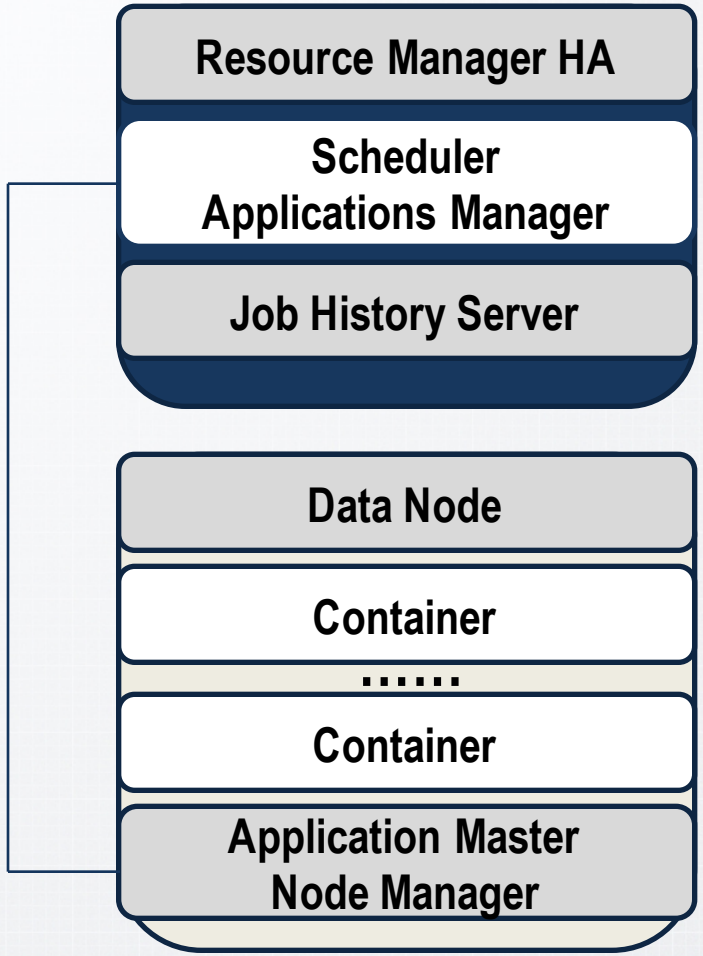
- An HDFS Administrator may have access to the data in encrypted form; but, he does not have access to the encryption keys.
- An HDFS user can only access the cipher text of files for which he has HDFS permissions.
- An HDFS user can only decrypt files based upon KMS ACLs associated to the EDEK.
- A Name Node does not have access to the key materials (i.e. DEKs and EZKeys). This prevents HDFS daemons from exposing sensitive data.

Performance Impacts



Role Based Access Control (RBAC)

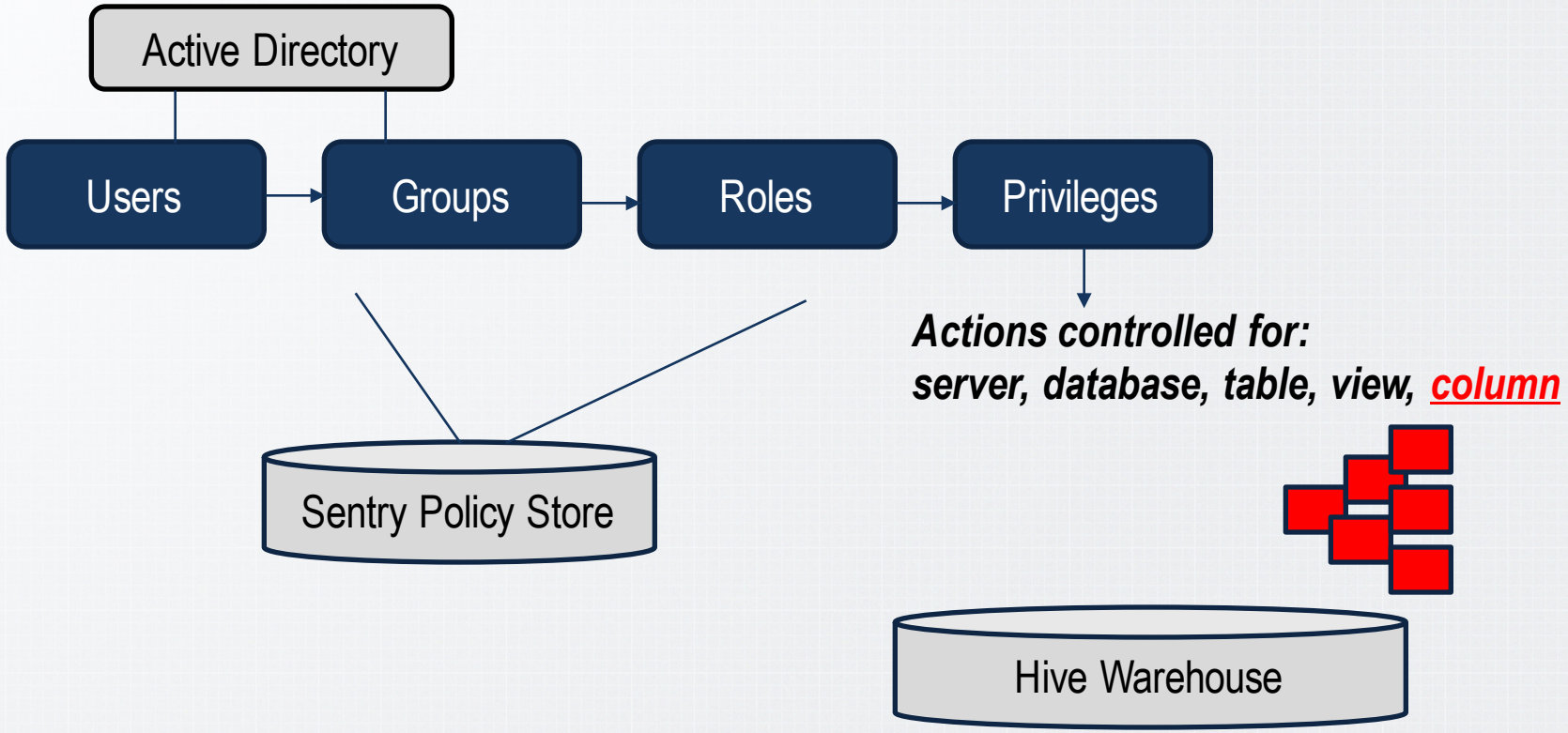
Job Isolation and Resource Control



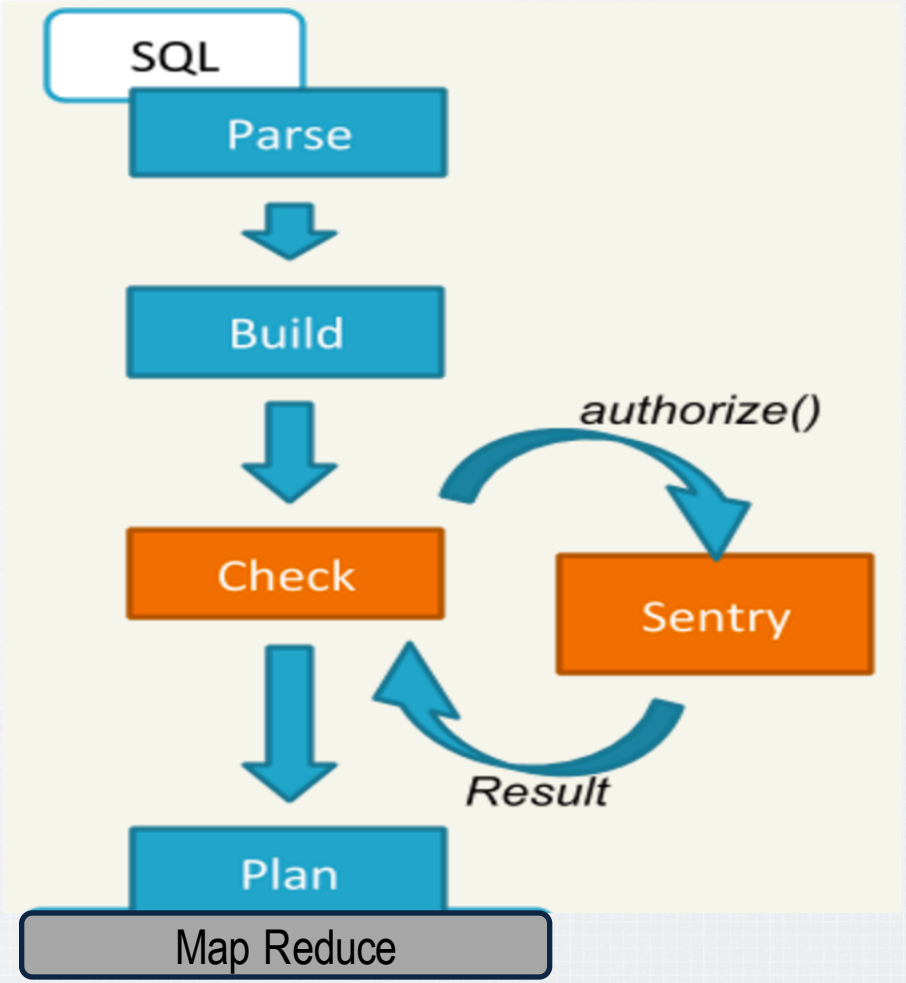
- Node Manager uses Kerberos authentication
- Also uses setuid when launching a YARN container to handle a job.
- Jobs run as the user who submitted the job and within isolated containers to avoid trespassing

  Impersonation
- Configured in yarn-site.xml

Apache Sentry



HiveServer2 Hook



Column Masking

Assign Sentry privileges to view

Java User-Defined Function (UDF)

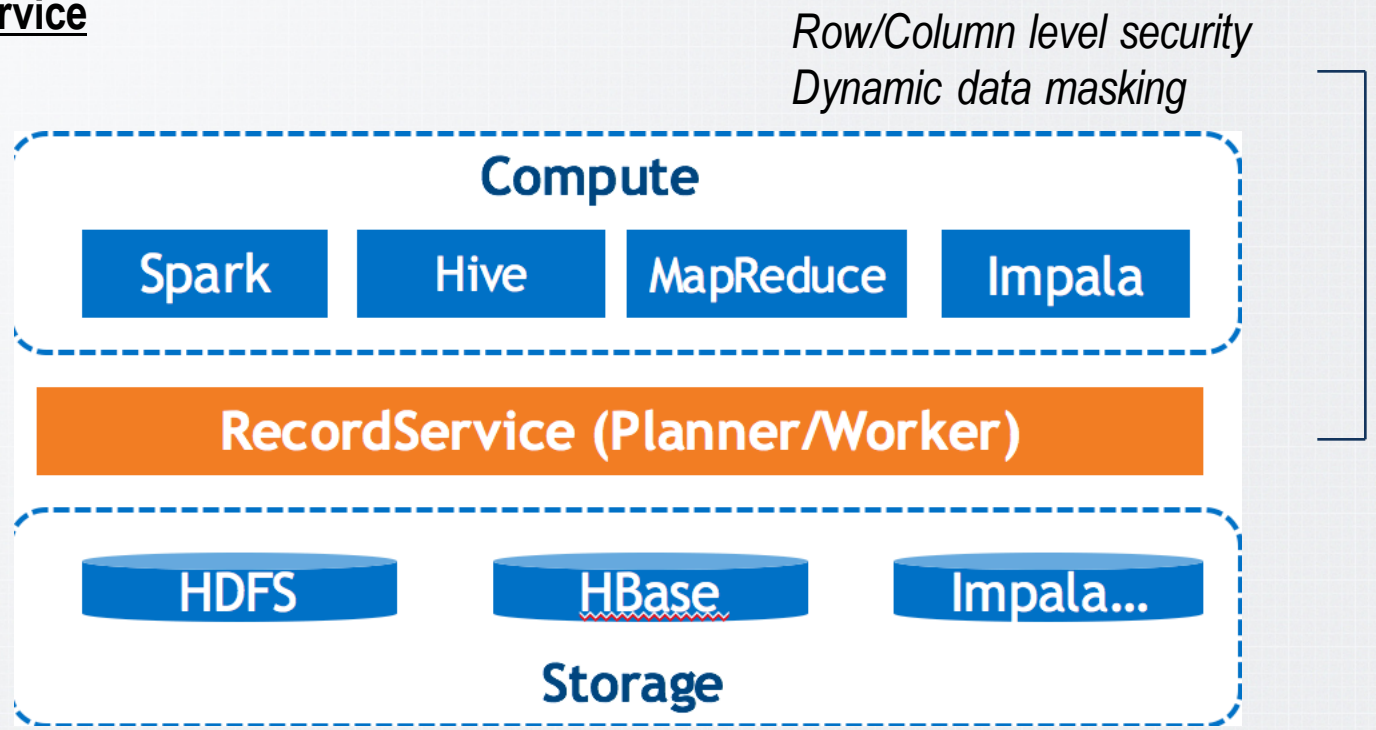
```
USE
ETL_STAGE_{source_database}

CREATE VIEW PII_MASKED_EXAMPLE
as
   SELECT mask_ccn_udf(credit_card_number) as
          ccn, name, balance, region
   FROM
   ETL_STAGE_VIEW_{source_database}.{Table}
   WHERE state = “VA”
```

Evolving Landscape

- Before column-level security was supported in Apache Sentry, views with access controlled by a Sentry policy were used

- **Cloudera Record Service**



Auditing

SIEM


CHALLENGERS
LEADERS
IBM
Splunk
LogRhythm
EMC (RSA)
HPE
Intel Security
Micro Focus
Trustwave
AlienVault
SolarWinds
Fortinet (AccelOps)
ManageEngine
EventTracker
BlackStratus
NICHE PLAYERS
VISIONARIES
ABILITY TO EXECUTE
COMPLETENESS OF VISION
As of August 2016

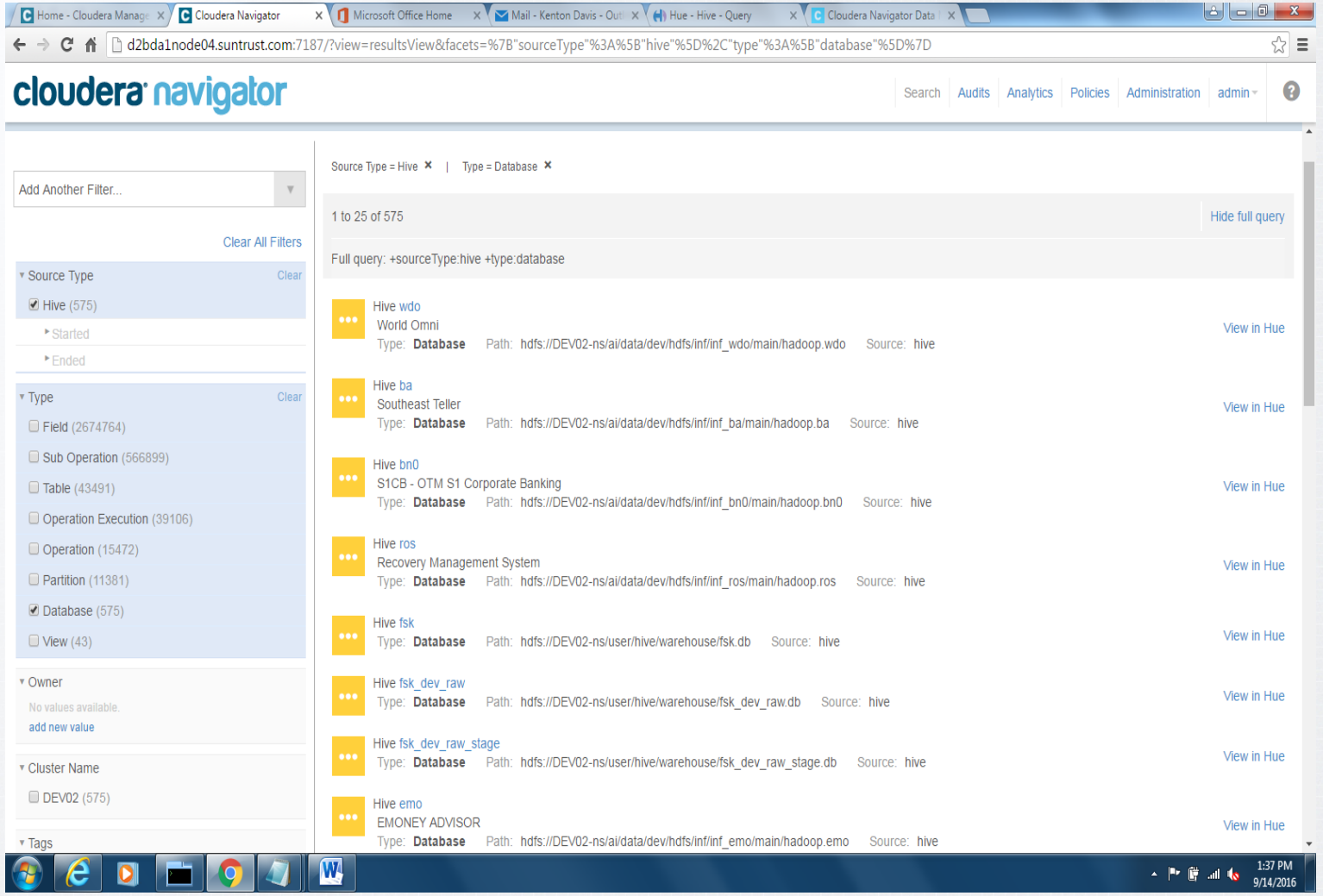
Splunk Monitoring

- Use Splunk TA forwarders on the BDA Nodes to capture Hadoop logs
- Slowly develop a profile of typical event data to identify suspicious behavior

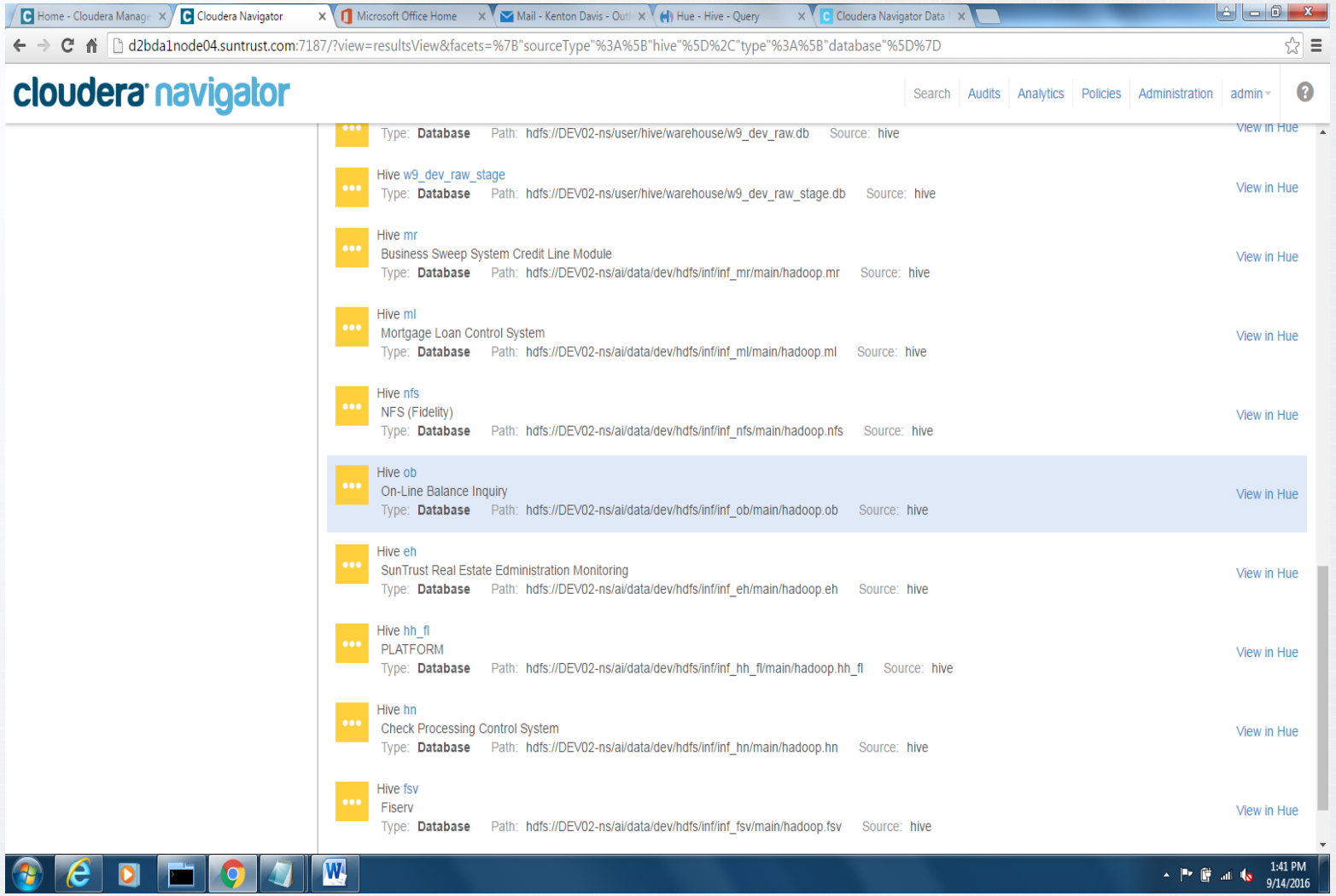
Compliance concerns addressed by Splunk

- PCI-DSS
- FISMA
- SOX

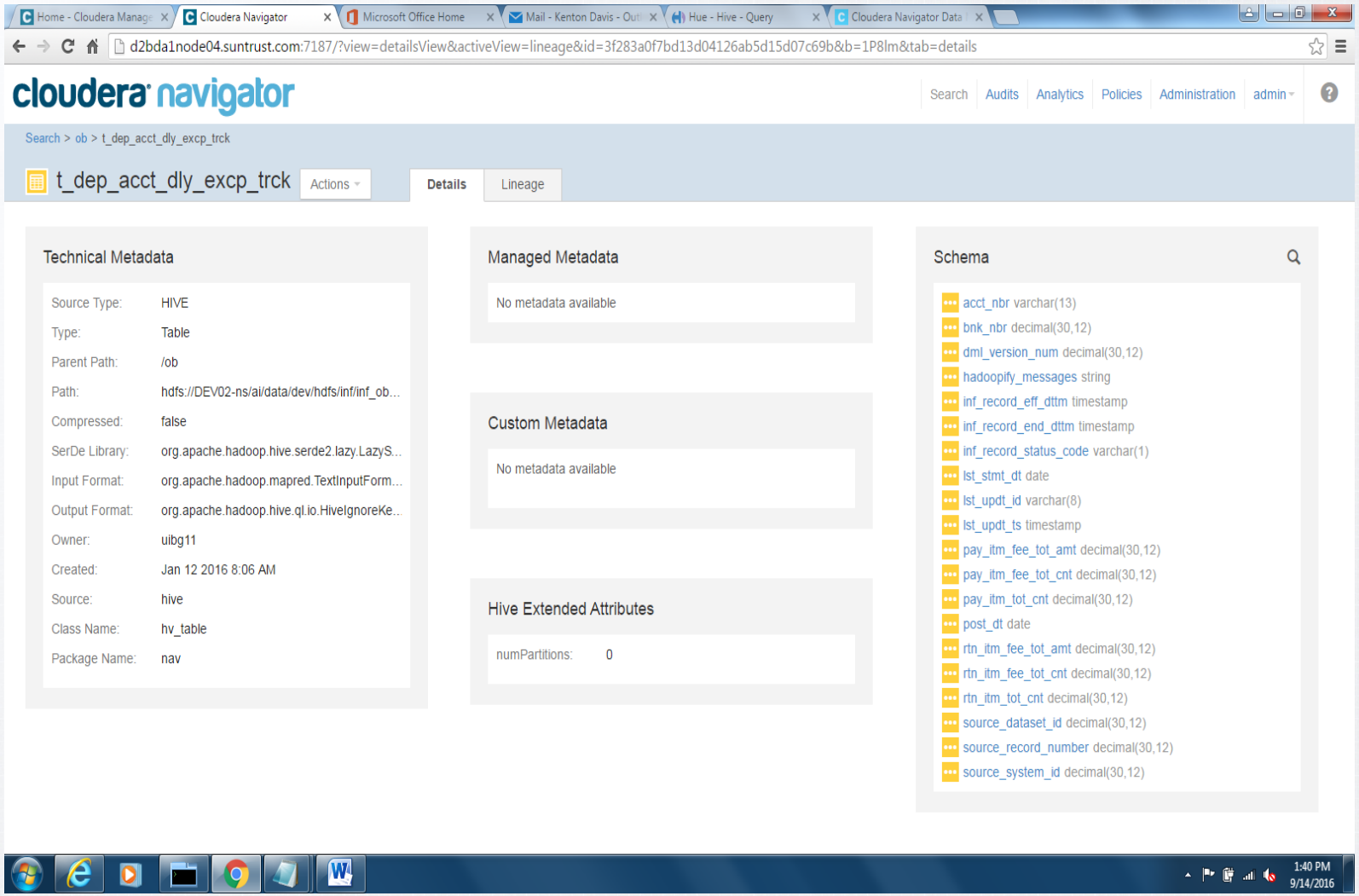
Lineage Tracking - #1



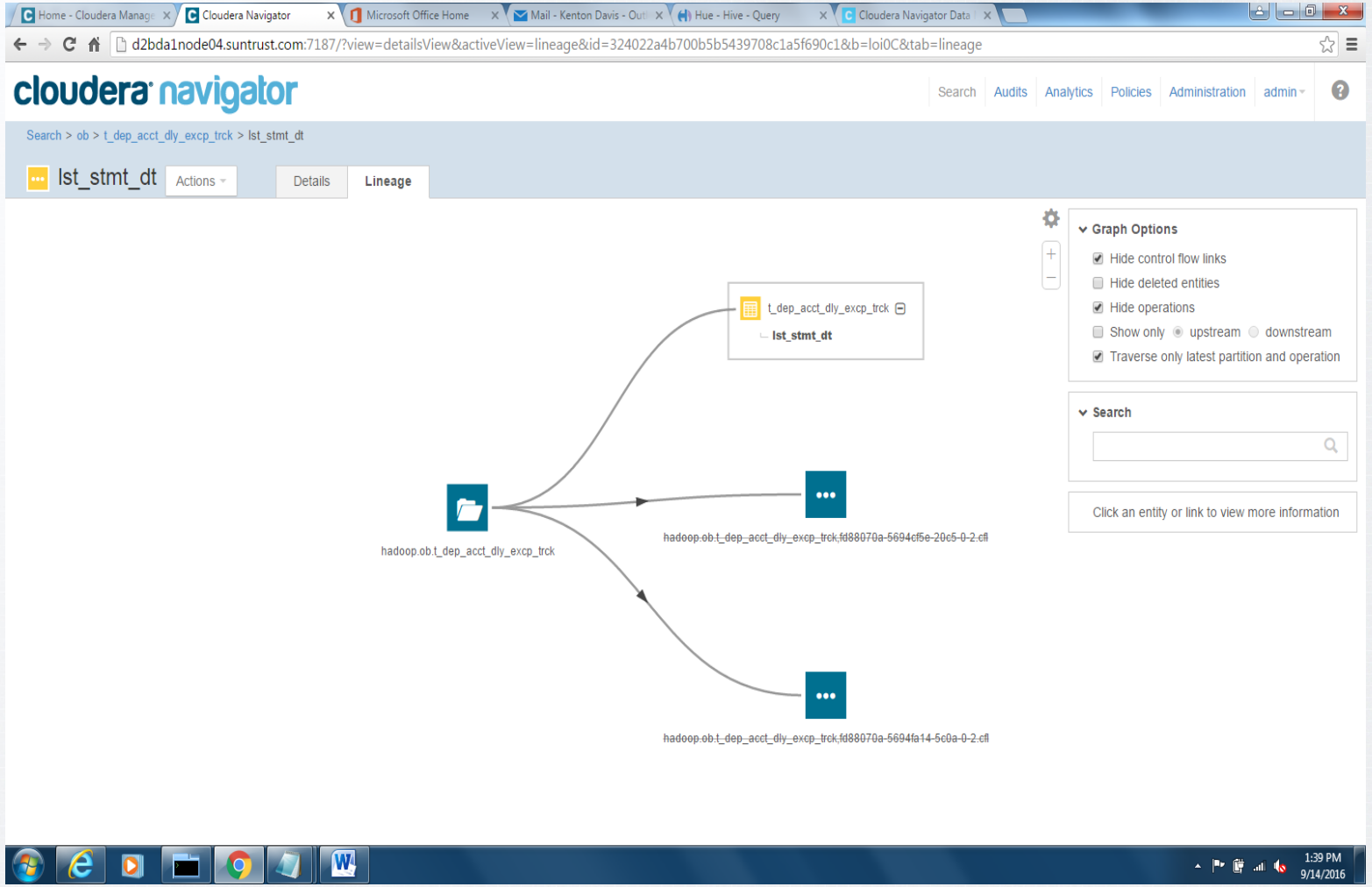
Lineage Tracking - #2



Lineage Tracking - #3



Lineage Tracking - #4



Challenges and Recommendations

Challenges during deployment

- AES-NI instruction sets were not present and AES-Generic support had to be manually loaded.
- Copying of data between clusters and from YARN Gateways into encryption zones using ETL/ELT products required some deployment considerations of where to situate the KMS.
- WAN Replication between Primary and DR sites invited concerns over how copying between encryption zones works.

Challenges during deployment

- The use of Centrify for authentication against Active Directory from Linux invites the need to wait for Centrify support for versions of OpenSSL to keep up to date with vulnerability issues found in CVE releases.

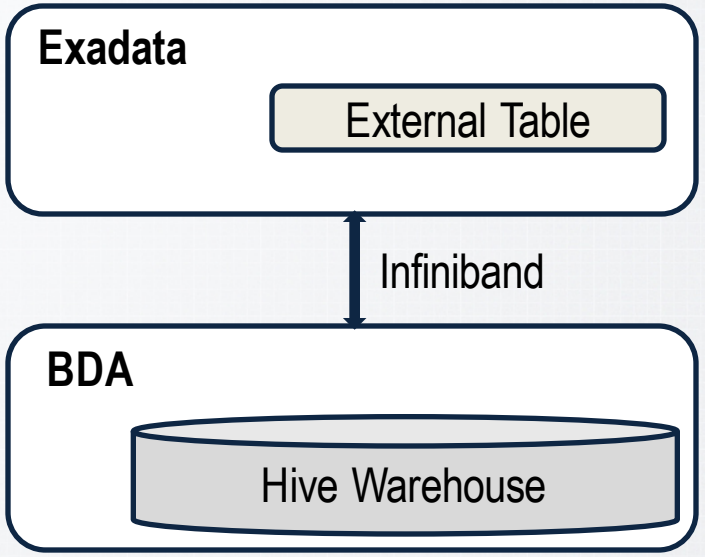
- Active Directory group mappings had to be redone for administrators to support separation of duties (Hive Admin?, Hadoop Admin?, Key Admin?)

- Is difficult to enforce all job submission through gateway nodes and to enforce */sbin/nologin* on cluster nodes.

Oracle Big Data SQL

Query Franchising

CREATE TABLE … ORGANIZATION EXTERNAL (TYPE oracle_hive);



Data Redaction to transform data on-the-fly (e.g. for credit-card masking)

DBMS_REDACT.ADD_POLICY

Virtual Private Database context predicates for row-level security

DBMS_RLS.ADD_POLICY

HOW TO MOVE TO THE CLOUD

Architect

Migrate

Integrate

Secure

Manage

www.biascorp.com/cloud

BIAS
Business Intelligent Application Solutions

ORACLE Platinum Partner


Q&A

Contact Us

Kenton Davis

Kenton.Davis@biascrop.com