



CLOUDERA

Cloudera Navigator

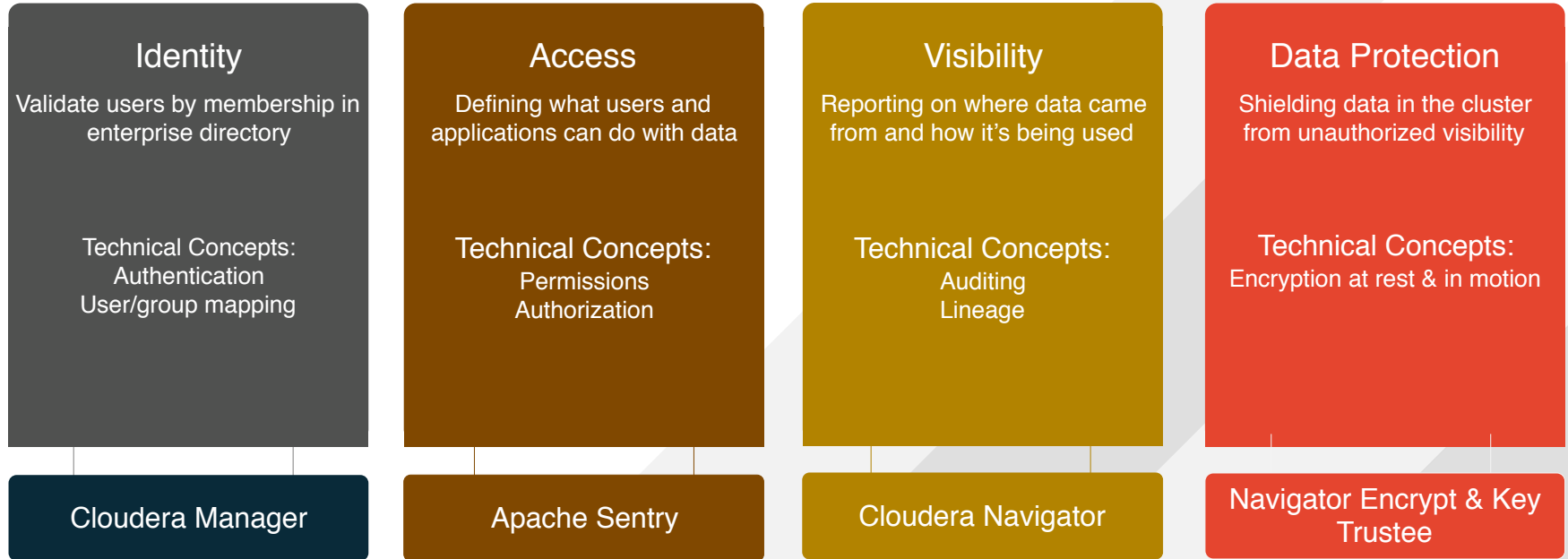
Enterprise Data Governance

Johannes Muselaers
Solutions Engineer Nordics
jmuselaers@cloudera.com
+46 72 588 1091

PDF of this presentation:

<https://github.com/jmuselaers/Presentations>

Cloudera Enterprise-Grade Security and Governance



CLOUDERA NAVIGATOR

Governance, stewardship, and discovery for big data built on machine learning and analytics



Trusted for production

- Deployed by hundreds of customers across multiple industries
- Over five years in production

Compliance-grade

- The only Hadoop distribution to pass PCI audit

Open and interoperable

- Integrated with leading partner solutions

Cloudera Navigator

- **Audit & Access Control**

- Maintain full audit history across Hue, Cloudera Manager, HDFS, Impala, HBase & Solr
- Ensuring appropriate permissions and reporting on data access for compliance

- **Discovery & Exploration**

- Finding out what data is available and what it looks like
- Tag datasets with metadata for pooling, knowledge sharing and common definitions

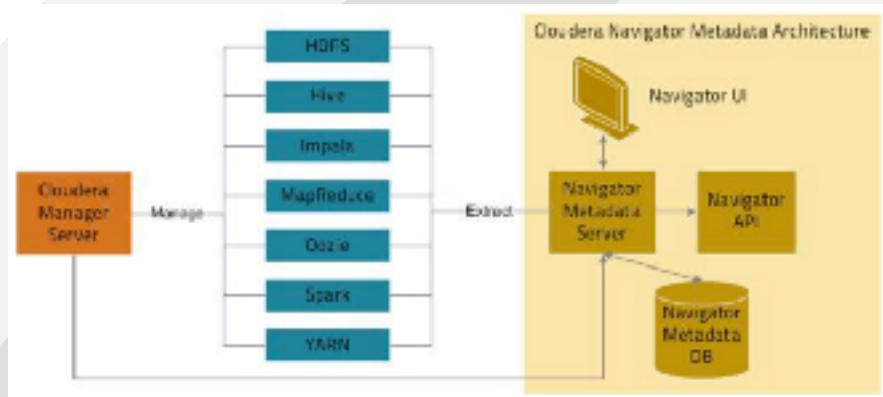
- **Lineage**

- Tracing data back to its original source
- Download and consume lineage data

- **Policy Engine**

- automated classification on arrival

- **SDK for integration with 3rd party tooling**



cloudera navigator

- Who (IP), What (operation), When (timestamp) across the stack
- Any HDFS file access
- Denied logins to Hue, CM, Navigator
- Query text from Impala, Hive, HBase, Solr, Sentry

- Who (IP), What (operation), When (timestamp) across the stack
- Any HDFS file access
- Denied logins to Hue, CM, Navigator
- Query text from Impala, Hive, HBase, Solr, Sentry

Lineage

cloudera navigator

Select > salesdata



salesdata

Actions

Details

Lineage



8-MB-CAT3



[View Full Details](#)

Source Type: HIVE
Type: Table
Path: hdfs://mosekty-lapcloud...
Compressed: true
SerDe Library: org.apache.hadoop.hive.serde...
Input Format: org.apache.hadoop.mapred...
Output Format: org.apache.hadoop.hive.d...
Owner: training
Created: Nov 10 2015 10:37 PM
Created: HIVE 1

Description: This is the raw data files history from March 2014 in New York. Obtained by monthly MIA feed.
Tags: salesdata, mla, salesdata, sensitive
month: March
retain_until: 2016-03-15
source: mynote
Steward: John Smith

AUTOMATED DATA CURATION

- Set up policies that classify data sets automatically upon ingest
- Add business glossary definitions and profiling information for faster self-service
- Automatically trigger data preparation, profiling, and data quality activities

Archive files older than 7 years

STATUS ✔ Enabled

Source Group: [source_group@cloudera.com](#)

Policy description: [source_group@cloudera.com](#)

Last Run On: Thursday, July 14, 2016, 10:00 AM

Last Modified: Friday, April 1, 2016, 12:00 PM

Last Modified By: admin

Schedule: Hourly

Start Time: Jan 27, 2017 9:30 AM

Run Times: Dec 01, 2024 9:00 AM

Interval: 1 month()

Metadata Assignments

Name	
Description	
Type	archive
Key-Value Pairs	

Retention Policies

Type	None
Target Path	archive

File Attributes

Modification	archive file
Owner	cloudera@cloudera.com

Archive files older than 7 years

STATUS ✔ Enabled

Source Group: [source_group@cloudera.com](#)

Policy description: [source_group@cloudera.com](#)

Last Run On: Thursday, May 20, 2016, 10:00 AM

Last Modified: Saturday, May 14, 2016, 12:00 PM

Last Modified By: admin

Schedule: On Change

Metadata Assignments

Name	
Description	This is the real source data history from March 2014 to April 2016. Created by monthly MLI feed.
Type	archive

Retention Policies

Type	None
Target Path	archive

File Attributes

Modification	archive file
Owner	cloudera@cloudera.com

← → ↺ 🏠

governance-svc-1.grandblue.com:7187/?view=detailsView&id=128897026b-1wFD01

☆ 🔌 📺 📁 🏠 🌐

📱 Apps

📄 Recent Documents

📁 DSWs

📁 NewAPI

📁 Remote Systems

🔑 Cloudera - Sign In

📁 NUC - CM

📁 Azqua CI - CM

📁 Gen44

📁 Hie

📁 Cloudera Manager

📁 Cloudera

cloudera NAVIGATOR

Search

Audits

Analytics

Policies

Administration

admin

customers_kudu

/user/hive/warehouse/customers_kudu

Details

Lineage

📁 Directory HDFS

Owner: impala
Parent Directory: warehouse
Group: hive
Permissions: rwxrwxrwl

1 Inputs

0 Outputs

May 2, 2018 4:18 PM Modified

▼ Technical Metadata

Source Type	HDFS
Parent Directory	warehouse
Path	/user/hive/warehouse/customers_kudu
Owner	impala
Group	hive
Permissions	rwxrwxrwl
Last Modified	May 2, 2018 4:18 PM
Created	May 2, 2018 4:18 PM
Source	HDFS-1
Classname	HDFS Entry
Package Name	new

▼ Directory Contents (0)

No matches found

► Inputs (1)

► Outputs (0)

Actions

View in Hue

Edit Metadata

Move

Move to Trash

Edit Metadata



Name

customers_kudu

Description

Managed Metadata

No Managed Metadata properties available.

Custom Metadata

Tags

sensitive x

Key-Value Pairs



Cancel

Save

customers_kudu

/user/hive/warehouse/customers_kudu

Actions

Details

Lineage



Directory
HDFS

Owner: [Impala](#)

Parent Directory: [warehouse](#)

Group: [hive](#)

Permissions: [rwxrwxrwt](#)

1

Inputs

0

Outputs

May 2, 2018 4:13 PM
Modified

Tags (1)

Edit

sensitive

Directory Contents (0)

No matches found

Technical Metadata

Source Type	HDFS
Parent Directory	warehouse
Path	/user/hive/warehouse/customers_kudu
Owner	impala
Group	hive
Permissions	rwxrwxrwt
Last Modified	May 2, 2018 4:13 PM
Created	May 2, 2018 4:13 PM
Owner	ADMIN

Inputs (1)

Outputs (0)

x



Add Filters Clear All Filters

 HDFS 1

Field

Directory 1


mpc | 9

[Find More Videos](#)

Cluster 1 1

Field Name: *Chlorophyll*


Share full access

 Hive a neighbor

Data Type: `string`

Parent Path [/default/top_10](#)


Source: HMF 1

[View in IJue](#) Hive a neighbor

Data Type: `string`

Parent Path [/default/salesdata](#)


Source: HIVF 1

[View in IJue](#) Hive a_neighbor

Data Type: `string`

Parent Path [/default/sales_by_region](#)

Source: HIVF 1

[View in IJue](#) Hive a neighbor

Data Type: string

Parent Path [/default/count_by_region](#)

Source: HIVF 1

[View in Hue](#) HDFS customers_cudr

Path: /user/warehouse/customers_kurdi

Owner Impulse

Group Inter:

Permissions [view](#) [edit](#)[View in French](#)

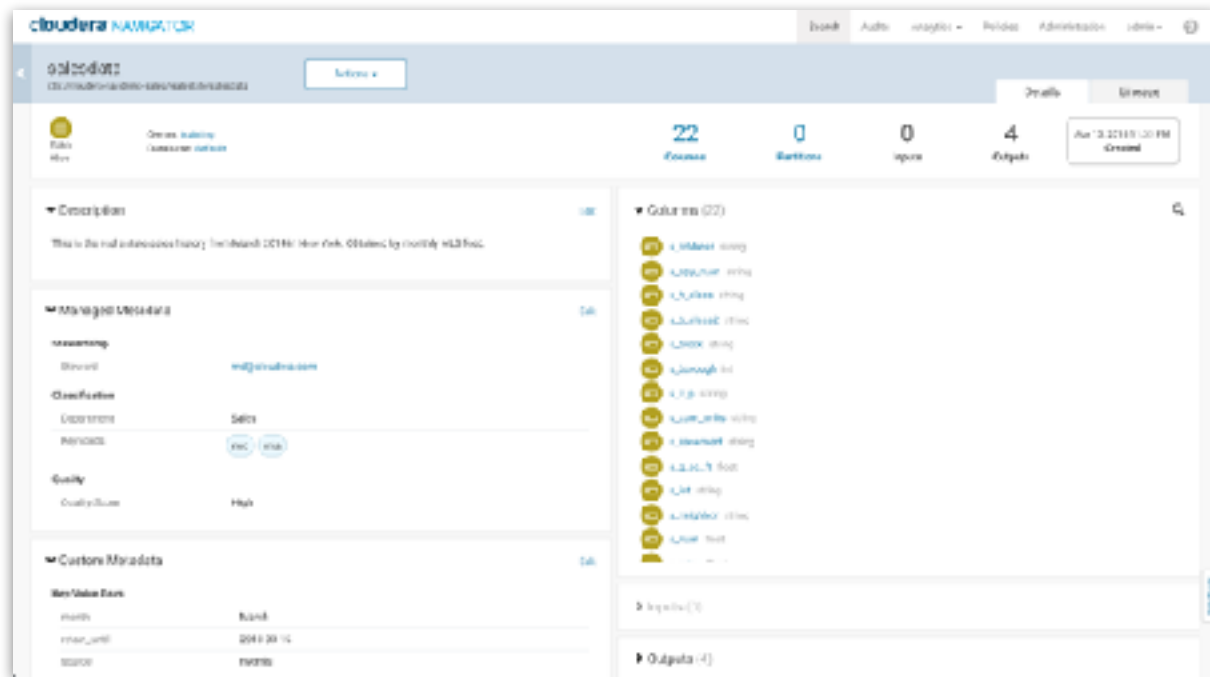
Last Modified: May 2, 2018 4:18 PM

Created: May 2, 2018 4:13 PM

Source: HDFS 1

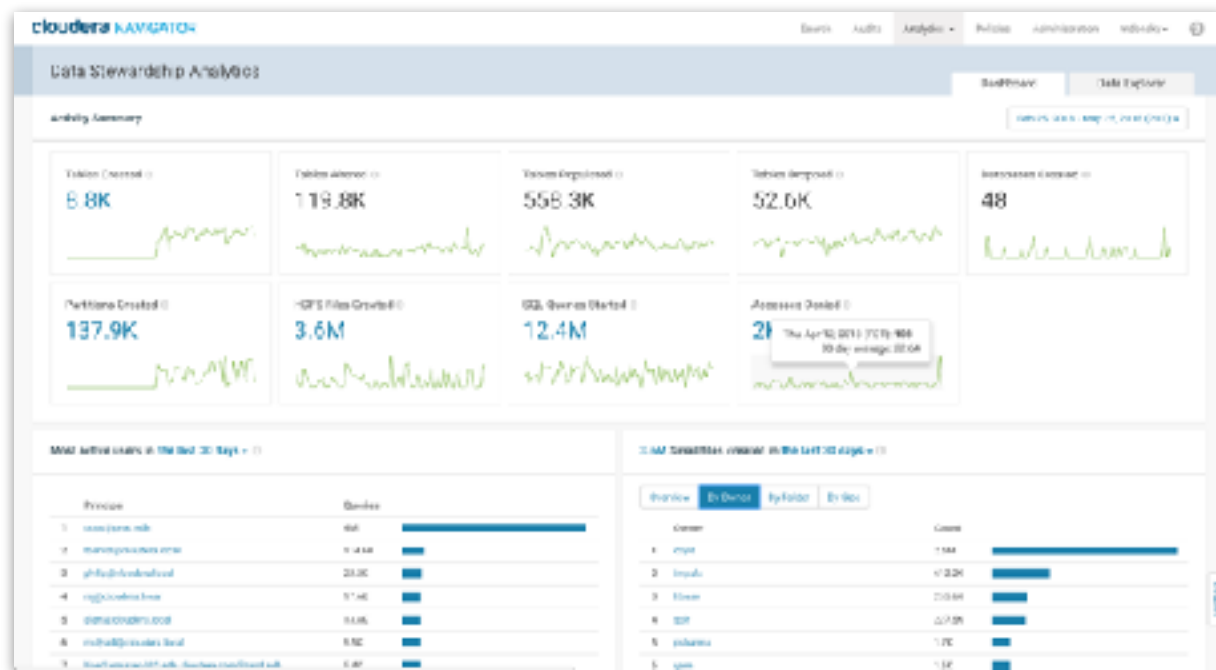
OUT-OF-THE-BOX SELF-SERVICE DISCOVERY

- Let business users collaborate on and classify data sets while leveraging centrally-curated classifications
- Leverage deep analytics on historical usage to empower users to find, trust, and use data sets on their own



STEWARDSHIP BUILT ON MACHINE LEARNING AND ANALYTICS

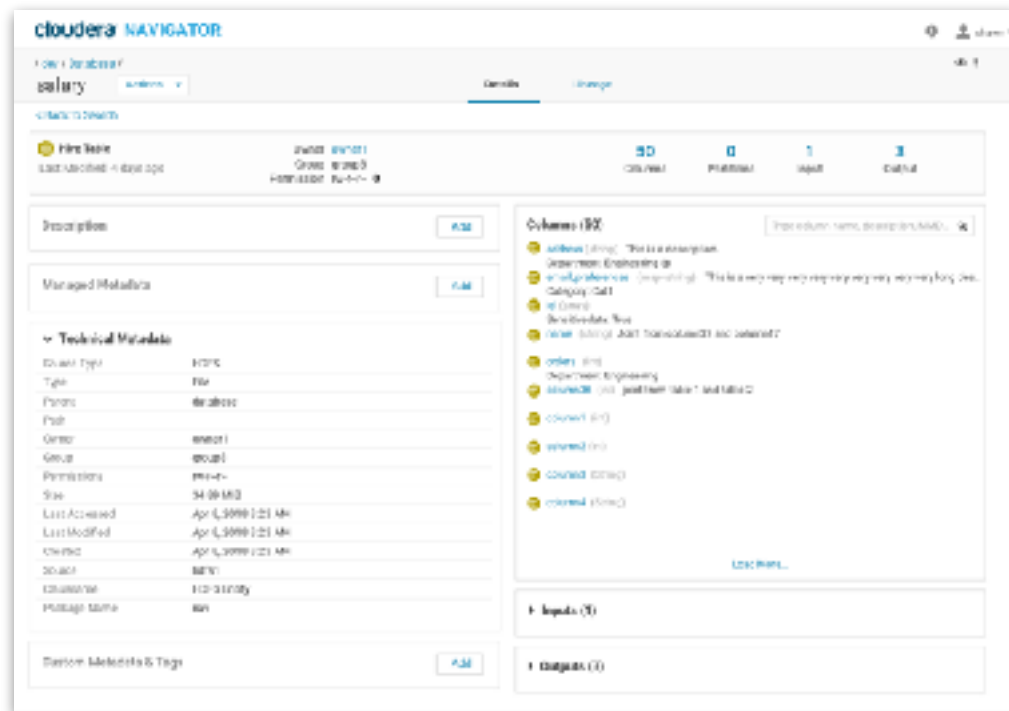
At-a-glance stewardship metrics



USE CASES | SELF-SERVICE DISCOVERY

Find, trust, and use data sets

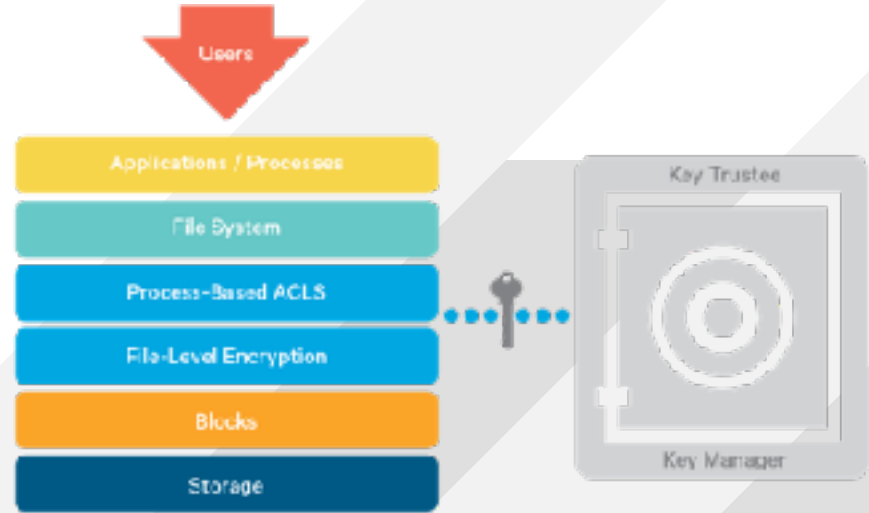
- **Find** data sets based on business context
- **Trust** data sets with insight based on ironclad governance
- **Use** data sets with analytics on on historical usage patterns



Navigator Encrypt

Transparent layer between
application and file system

Compliance-Ready
Massively Scalable
High Performance: Optimized for
Intel
Separation of Duties
Key Management with Navigator
Key Trustee



Latest releases

https://www.cloudera.com/documentation/enterprise/release-notes/topics/cn_rn_new_features.html

Roadmap



THANK YOU

CLOUDERA

<http://sdx-nav-demo-1.gce.cloudera.com:7187/login.html>