CLOUDERA

Open Source Data Management for Car-to-Cloud

Michael Ger
Managing Director, Manufacturing & Automotive
Cloudera

GENIVI®

CLOUDERA AT A GLANCE

One stop shop for Big Data management for analytics

Unified open source architecture

Hybrid and multi-cloud

© Cloudera, Inc. All rights reserved.

CLOUDERA ENTERPRISE DATA WAREHOUSE

Any Data. Anywhere. From Edge to AI

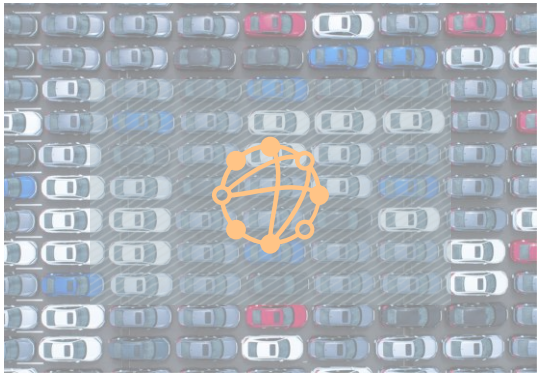
Traditional Data Warehouse Optimization

- BI Reporting
- Data Warehouse Offload
- ETL Optimization

Operations & Events Data Warehouse

- IOT Data (Sensors, Logs, etc.)
- Connected Vehicles
- Connected Factories (Industry 4.0)
- Etc.

Research & Discovery Data Warehouse



- Unstructured Data (Images, Video, etc.)
- Intelligent Search
- Autonomous Vehicle Research

© 2019 Cloudera, Inc. All rights reserved.

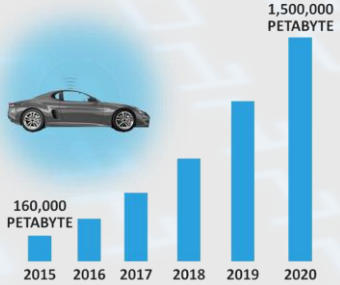
DID YOU KNOW?

10 of the top 10 Automotive OEMs
Trust Cloudera for Big Data-Enabled Digital Transformations

© 2019 Cloudera, Inc. All rights reserved.

DATA TRENDS

Connected Vehicle Data



10X by 2020*

* Source: Cowen and Company, Gartner

Autonomous Vehicle Data

TERABYTES

per vehicle, per day

Industrial Internet Data

2X faster than any other data source

* Source: Wikibon

Real-Time Data

1.5X faster than traditional data

* Source: IDC

© Cloudera, Inc. All rights reserved.

CAR TO CLOUD – DATA MANAGEMENT ON 5 LEVELS

THE EDGE

MANAGING THE EDGE

ENTERPRISE FLOW/STREAM ANALYTICS

EDGE TO AI

CONNECTED COMMUNITIES

© 2019 Cloudera, Inc. All rights reserved.

CAR TO CLOUD – DATA MANAGEMENT ON 5 LEVELS

THE EDGE

MANAGING THE EDGE

ENTERPRISE FLOW/STREAM ANALYTICS

EDGE TO AI

CONNECTED COMMUNITIES

© 2019 Cloudera, Inc. All rights reserved.

EDGE DATA COLLECTION - MINIFI AGENTS

- Extremely small footprint
- Java and C++ agents
  - JRHEL/CentOS, Debian/Ubuntu, Android*
- Key data management functions
  - Filtering, buffering, guaranteed delivery, prioritized queuing
- Secure
  - Encryption, certificate-based authentication
  - Data tagging and provenance
- Execute Machine Learning (ML) models including Tensorflow

© Cloudera, Inc. All rights reserved.

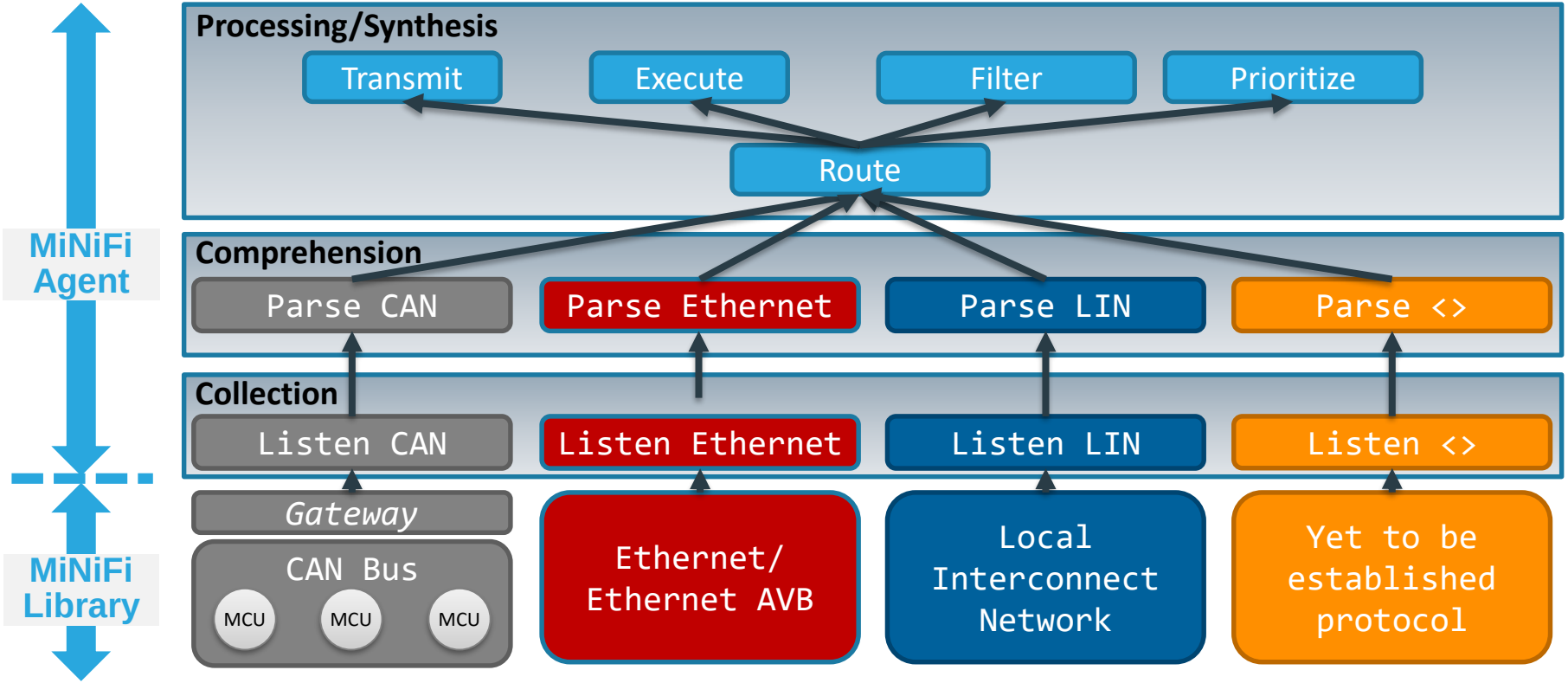
GOING SMALLER – FROM NIFI TO MINIFI

Supporting the Need for Smaller Footprints on the Edge

Application	Memory Footprint	Characteristics
MiNiFi Agent (Java)	10s to 100s of MBs	• Feature parity and reuse of core NiFi libraries
MiNiFi Agent (C++)	100s kBs to MBs	• Write once**, run anywhere. • Adaptable to lower level interfaces
MiNiFi Library	10s kBs	• Write n-many times, embed, run anywhere • Language libraries to support tagging, FlowFile format, Site to Site protocol and provenance generation without a full processing framework • Customizable to environments without disk storage or where threading is prohibitive • Language SDKs, Mobile Platforms

© Cloudera, Inc. All rights reserved.

MINIFI IN CONNECTED CAR



© Cloudera, Inc. All rights reserved.

CAR TO CLOUD – DATA MANAGEMENT ON 5 LEVELS

THE EDGE

MANAGING THE EDGE

ENTERPRISE FLOW/STREAM ANALYTICS

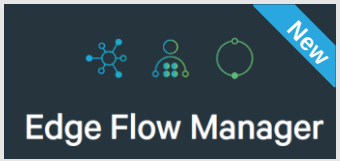
EDGE TO AI

CONNECTED COMMUNITIES

© 2019 Cloudera, Inc. All rights reserved.

CLOUDERA EDGE MANAGEMENT (CEM)

Edge Management Hub



Edge Agent Management

- **Central management** of agents
- **Collect data** from edge device
- **Push intelligence** to edge

Data

Intelligence

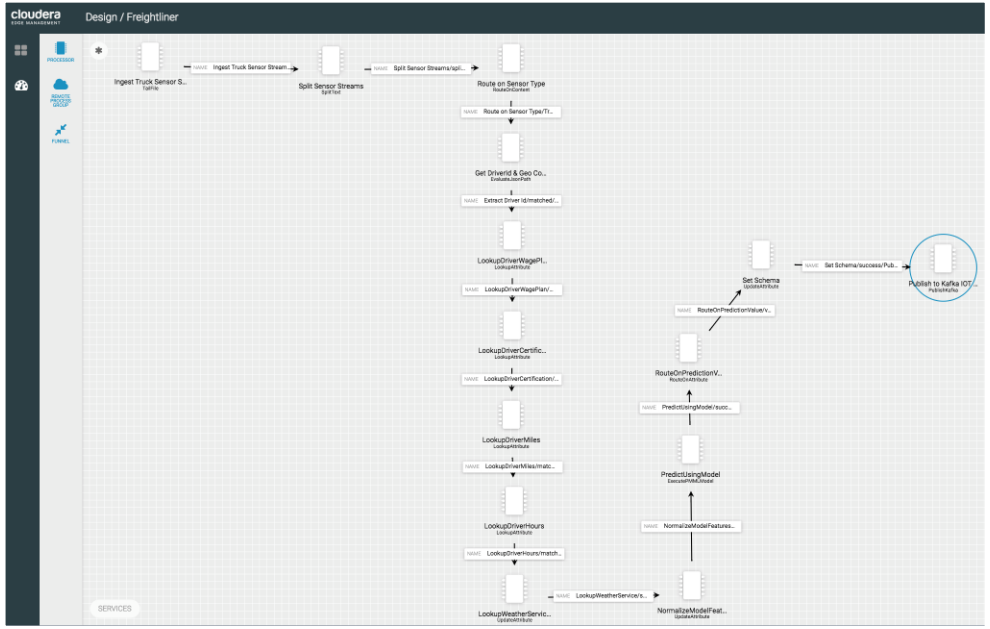
Edge Agents

- **Data Collection and Processing** at the Edge
- **Small, lightweight footprint**

© Cloudera, Inc. All rights reserved.

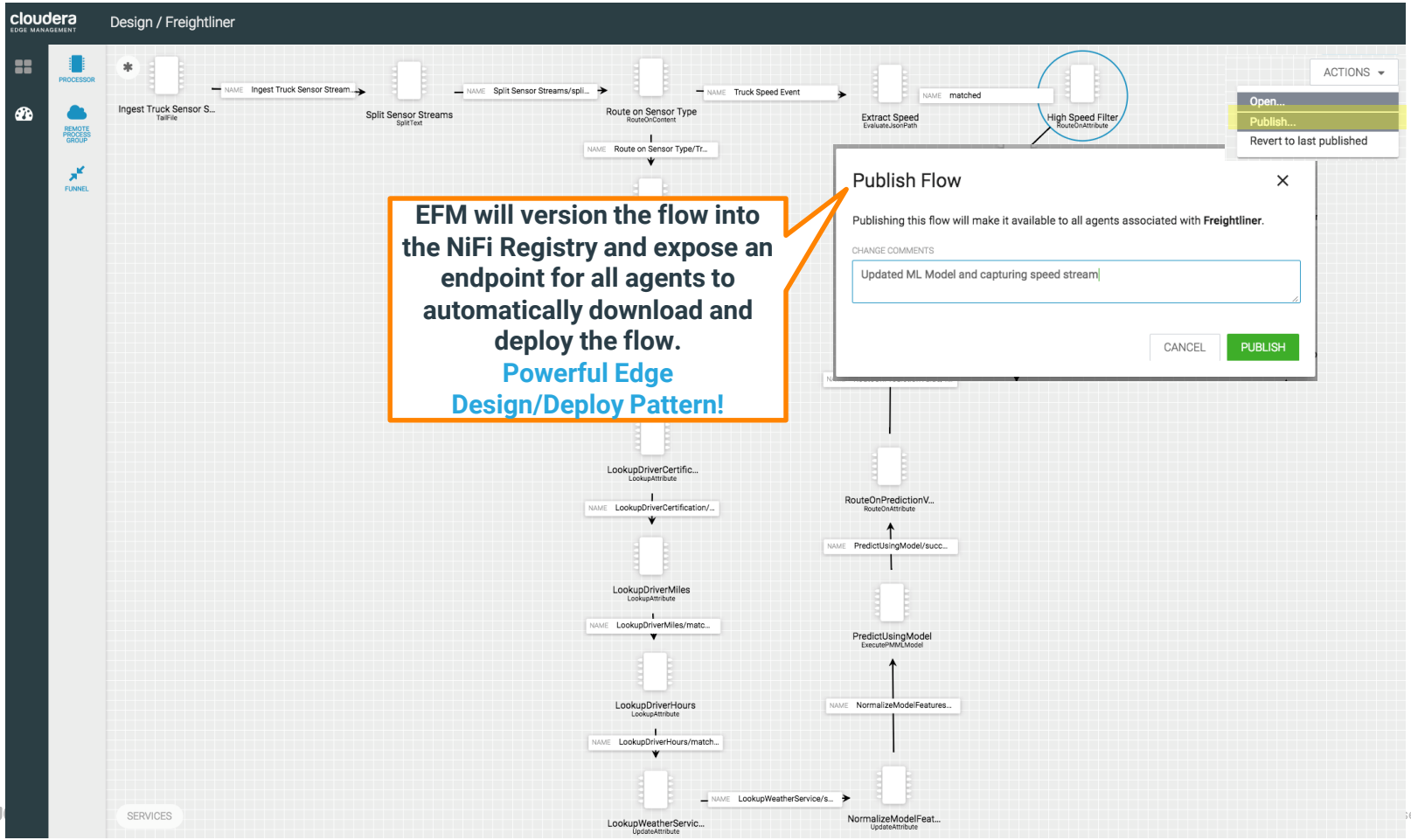
EDGE FLOW MANAGER

- **Edge Management Hub**
- Graphical user interface to develop and deploy flows to edge
- Monitor thousands of edge agents
- Deploy updated Machine Learning (ML) models to edge agents
- Integration with NiFi Registry



© 2019 Cloudera, Inc. All rights reserved.

INTELLIGENCE TO THE EDGE – PUSH ML MODELS TO EDGE



CAR TO CLOUD – DATA MANAGEMENT ON 5 LEVELS

THE EDGE

MANAGING THE EDGE

ENTERPRISE FLOW/STREAM ANALYTICS

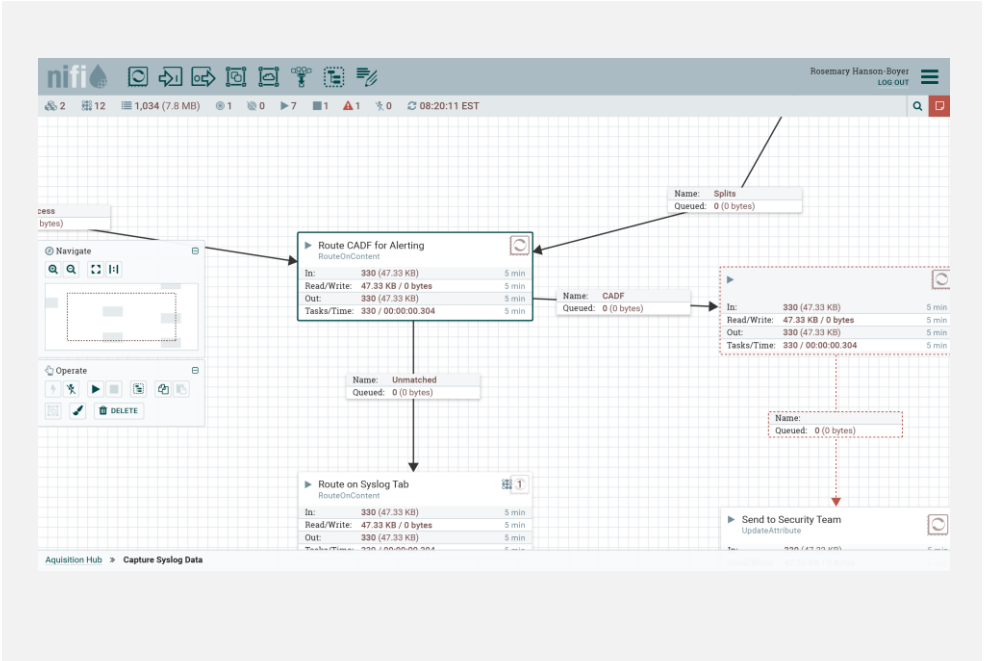
EDGE TO AI

CONNECTED COMMUNITIES

© 2019 Cloudera, Inc. All rights reserved.

FLOW MANAGEMENT VIA NIFI

- Based on Apache NiFi
- Highly configurable flow creation
- Web-based user interface
- 300+ prebuilt processors
- Secure - fine grained encryption
- Data provenance
- Guaranteed delivery
  - Buffer data during system interruptions
- Designed for extensibility
- NiFi Registry

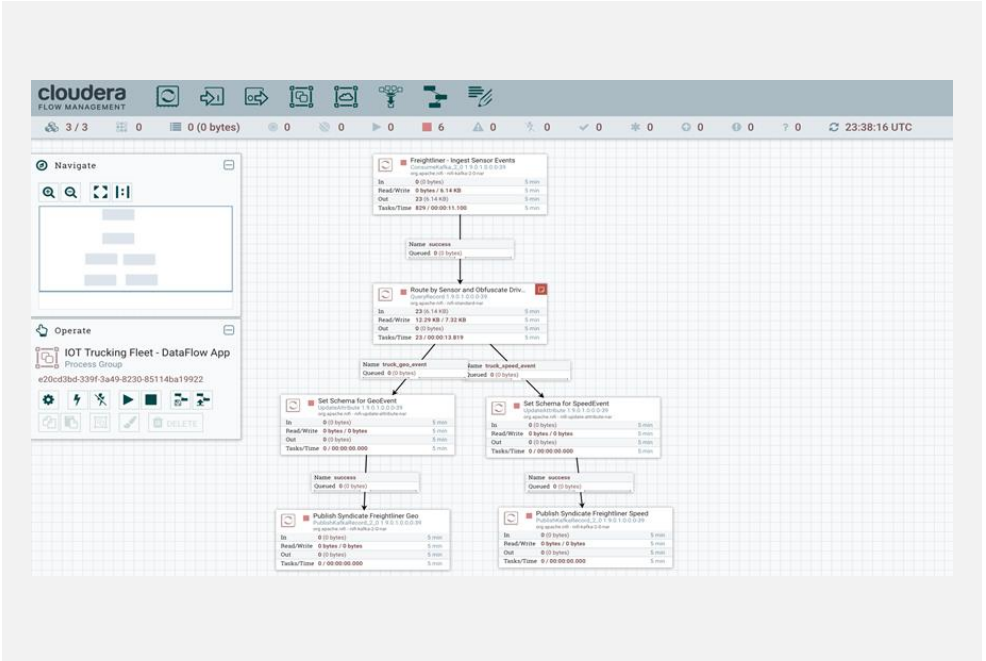


© 2019 Cloudera, Inc. All rights reserved.

WHAT CAN BE DONE WITH NIFI PROCESSORS?

300+ Pre-Built Processors

- **Ingestion**: connectors to read/write data from/to several data sources
  - **Protocols:** HTTP (S), AMQP, MQTT, UDP, TCP, CEF, JMS, (S) FTP, AWSIoT, Raw Socket Protocol
  - **Brokers:** Kafka, JMS, AMQP, MQTT etc.
  - **Databases:** JDBC, MongoDB, HBase, Cassandra etc.
- **Extraction** (XML, JSON, Regex, Grok etc.)
- **Transformation**
  - Format conversion (JSON to Avro, CSV to ORC etc.)
  - Compression/decompression, Merge, Split, encryption etc.
- **Data enrichment**
  - Attribute, content, rules etc.
- **Routing**
  - Priority, dynamic/static, based on content or metadata etc.



© 2019 Cloudera, Inc. All rights reserved.

300+ PROCESSORS FOR DEEPER ECOSYSTEM INTEGRATION

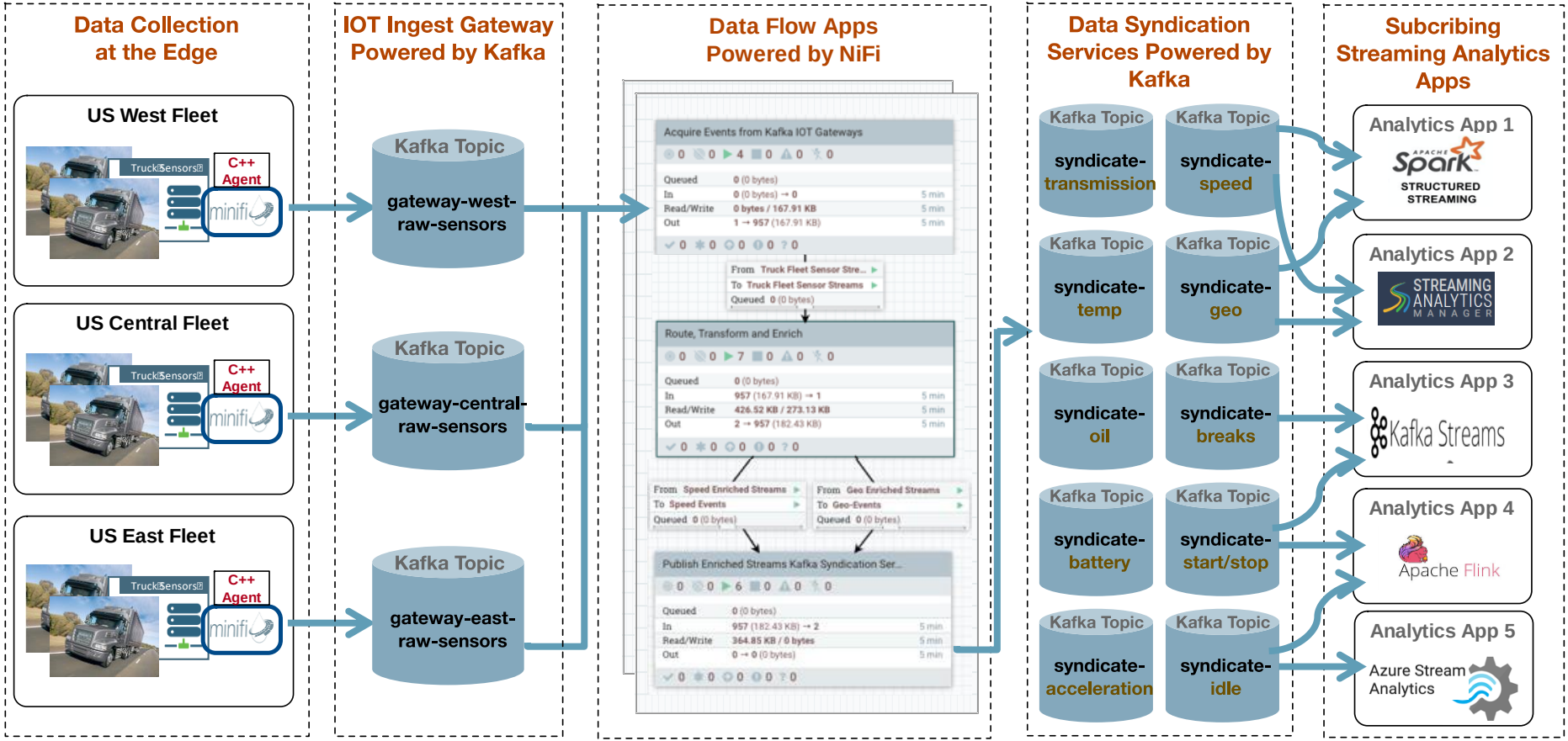
FTP
SFTP
HL7
UDP
XML
HTTP
WebSocket
Email
HTML
Image
Syslog
AMQP

Hash | Encrypt | GeoEnrich
Merge | Tail | Scan
Extract | Evaluate | Replace
Duplicate | Execute | Translate
Split | Fetch | Convert

Parse Records | Convert Records
Route Text | Distribute Load
Route Content | Generate Table Fetch
Route Context | Jolt Transform JSON
Control Rate | Prioritized Delivery

All Apache project logos are trademarks of the ASF and the respective projects.
© 2019 Cloudera, Inc. All rights reserved.

STREAMING ANALYTICS REFERENCE ARCHITECTURE



EXAMPLE - LAS VEGAS CONNECTED VEHICLE PILOT

City Fleet Vehicle In Operation

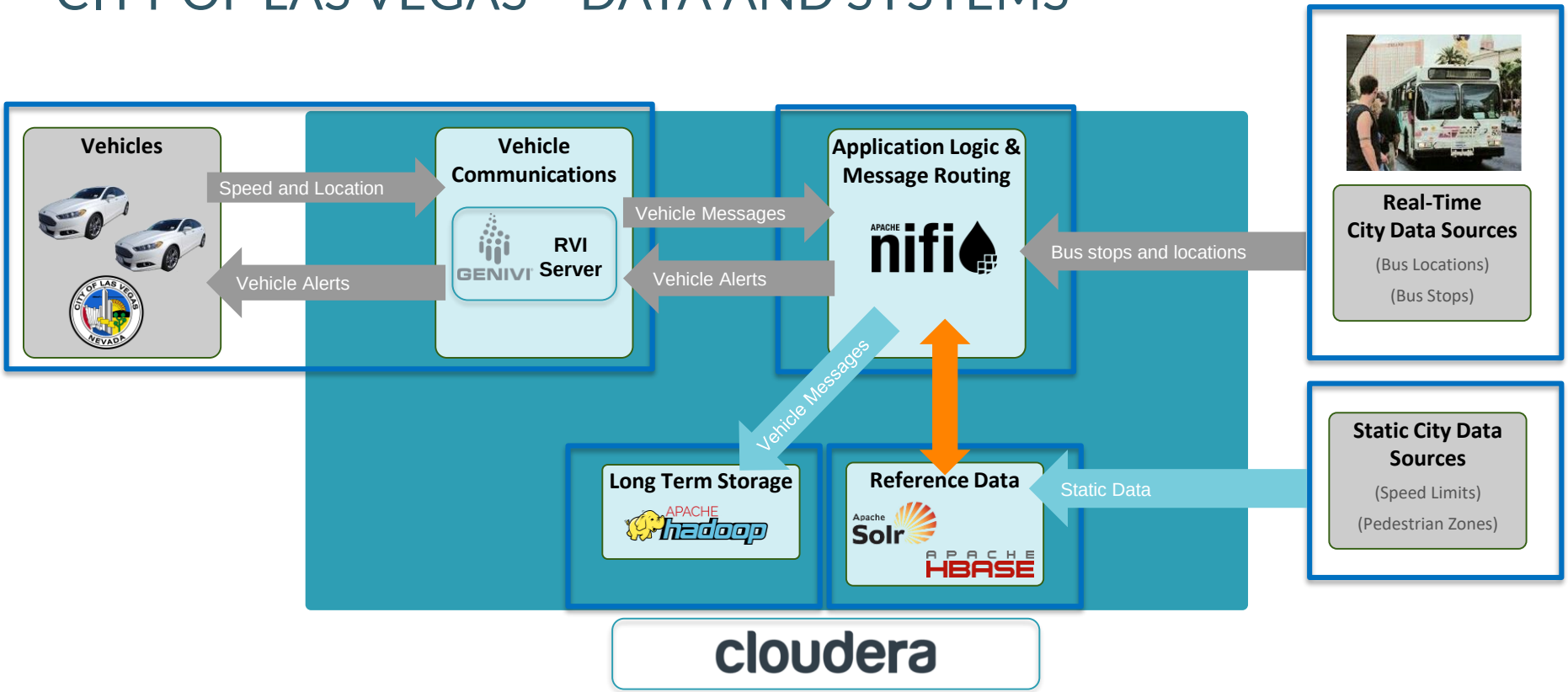
Monitor Vehicle Conditions

Approaching Active Bus Stop?

Issue Alert to In-Vehicle Display Unit

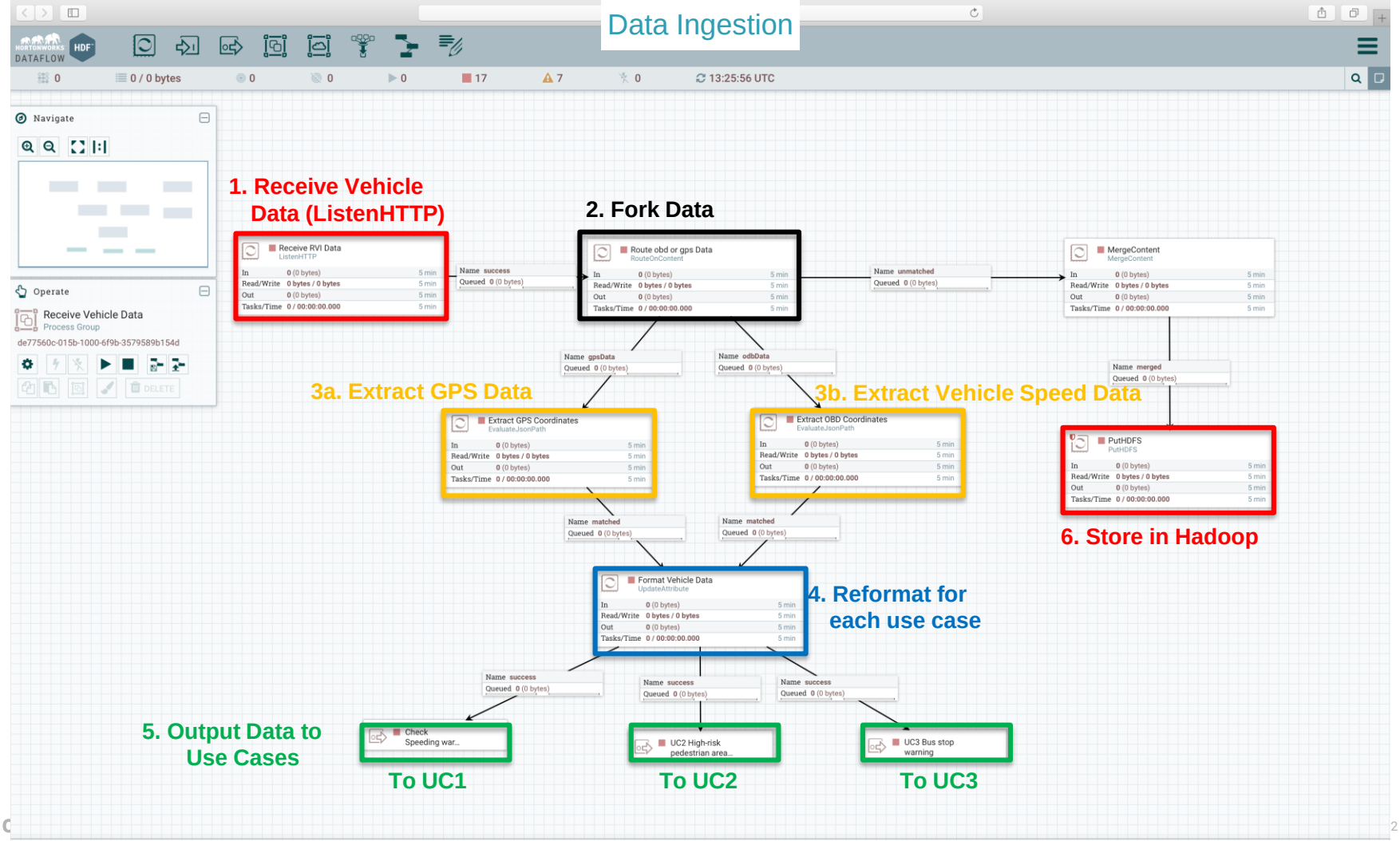
© 2019 Cloudera, Inc. All rights reserved.

CITY OF LAS VEGAS – DATA AND SYSTEMS

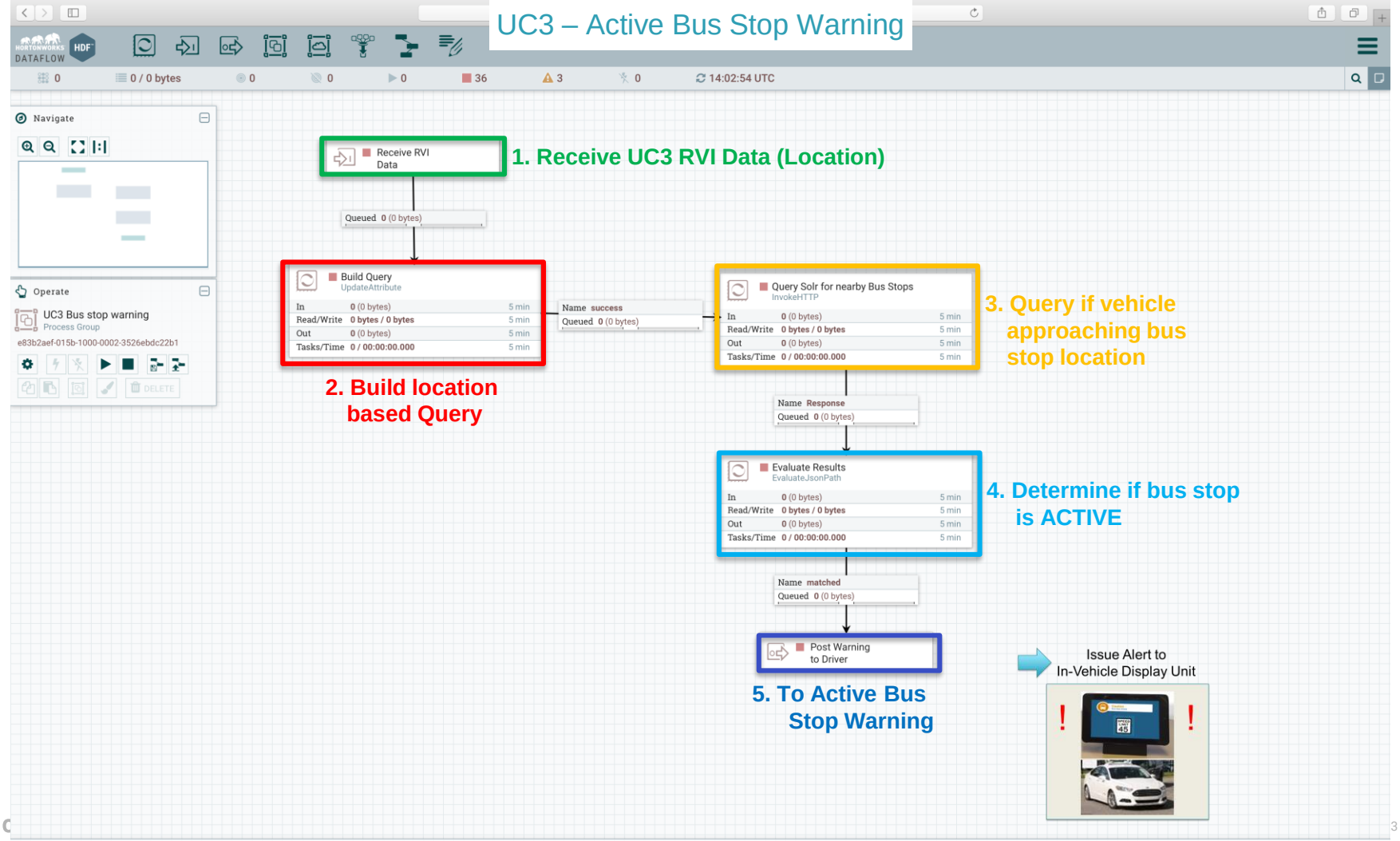


© Cloudera, Inc. All rights reserved.

Data Ingestion



UC3 – Active Bus Stop Warning



CAR TO CLOUD – DATA MANAGEMENT ON 5 LEVELS

THE EDGE

MANAGING THE EDGE

ENTERPRISE FLOW/STREAM ANALYTICS

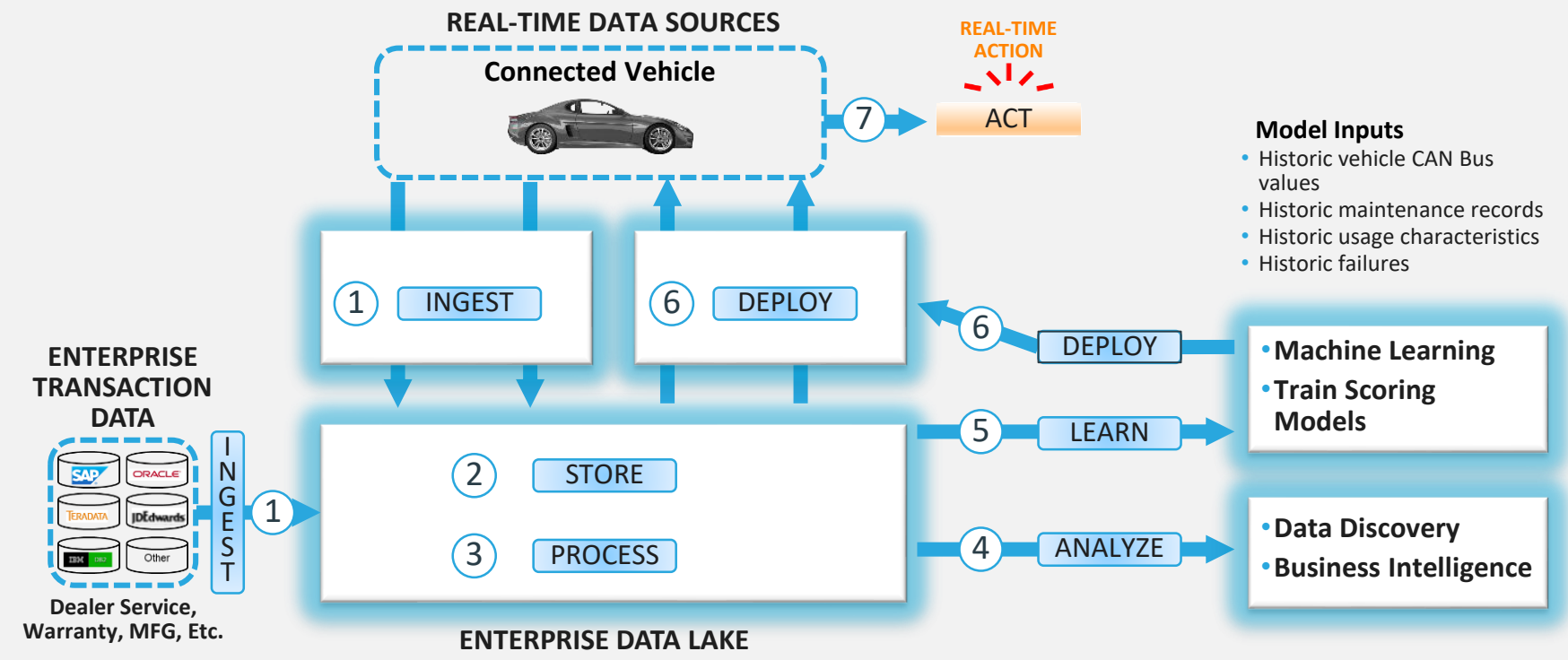
EDGE TO AI

CONNECTED COMMUNITIES

© 2019 Cloudera, Inc. All rights reserved.

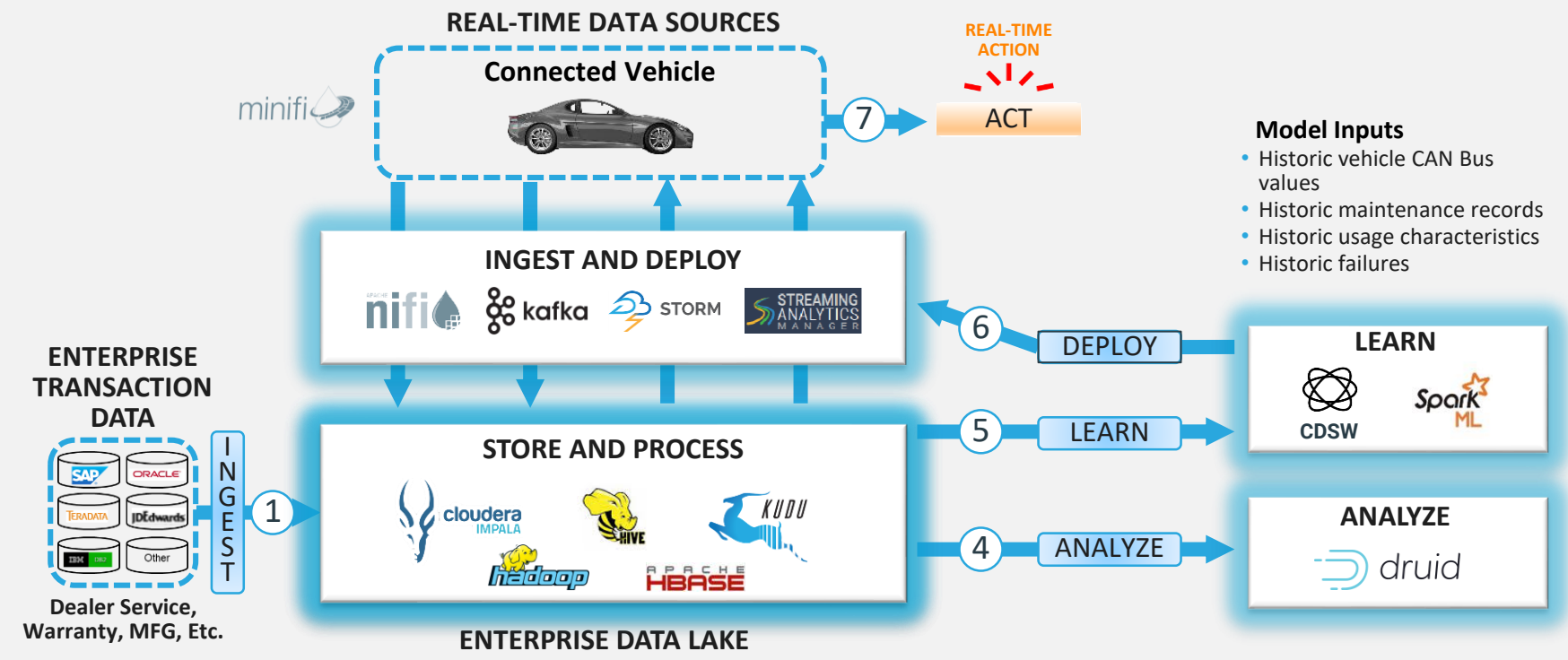
CONNECTED VEHICLE ANALYTICS LIFECYCLE

Example: Vehicle Predictive Maintenance



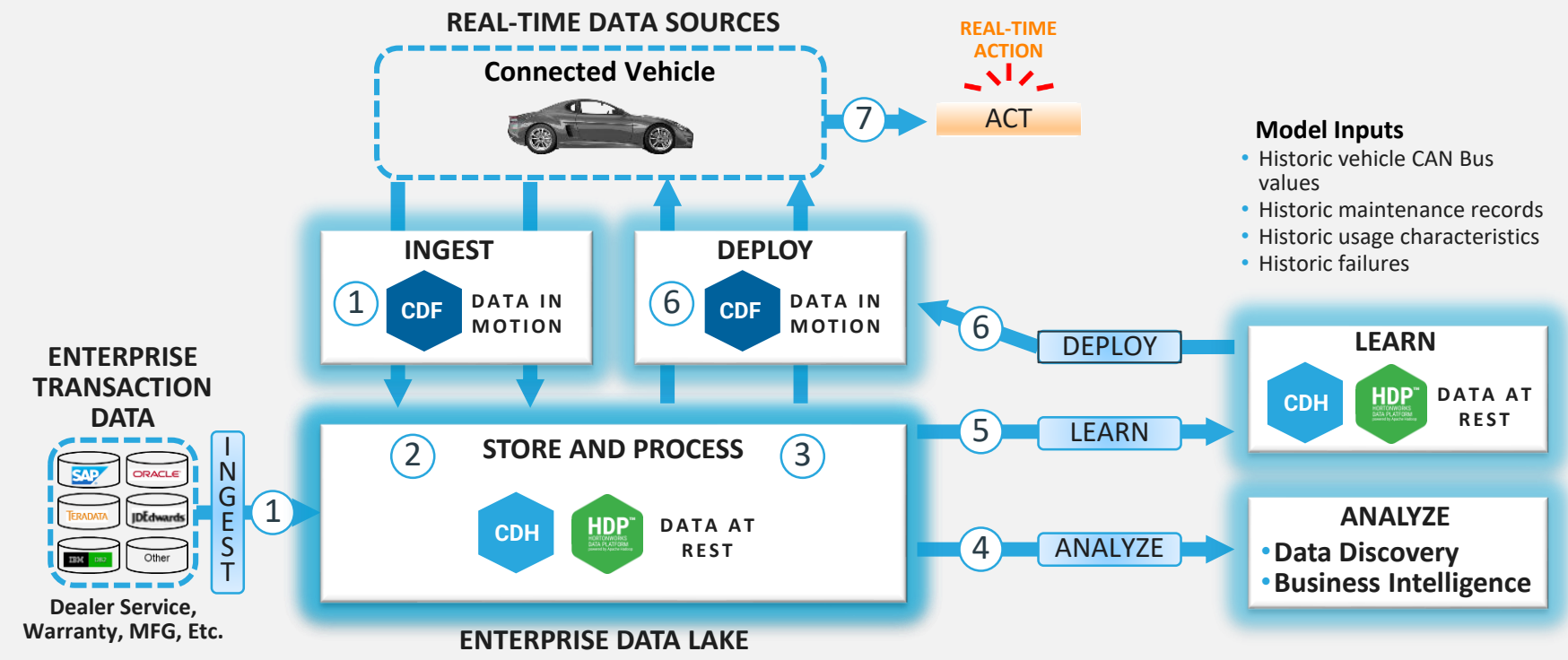
© 2019 Cloudera, Inc. All rights reserved.

SOLUTION COMPONENT DETAIL



© 2019 Cloudera, Inc. All rights reserved.

CLOUDERA FOR CONNECTED VEHICLE INNOVATION



© 2019 Cloudera, Inc. All rights reserved.

CAR TO CLOUD – DATA MANAGEMENT ON 5 LEVELS

THE EDGE

MANAGING THE EDGE

ENTERPRISE FLOW/STREAM ANALYTICS

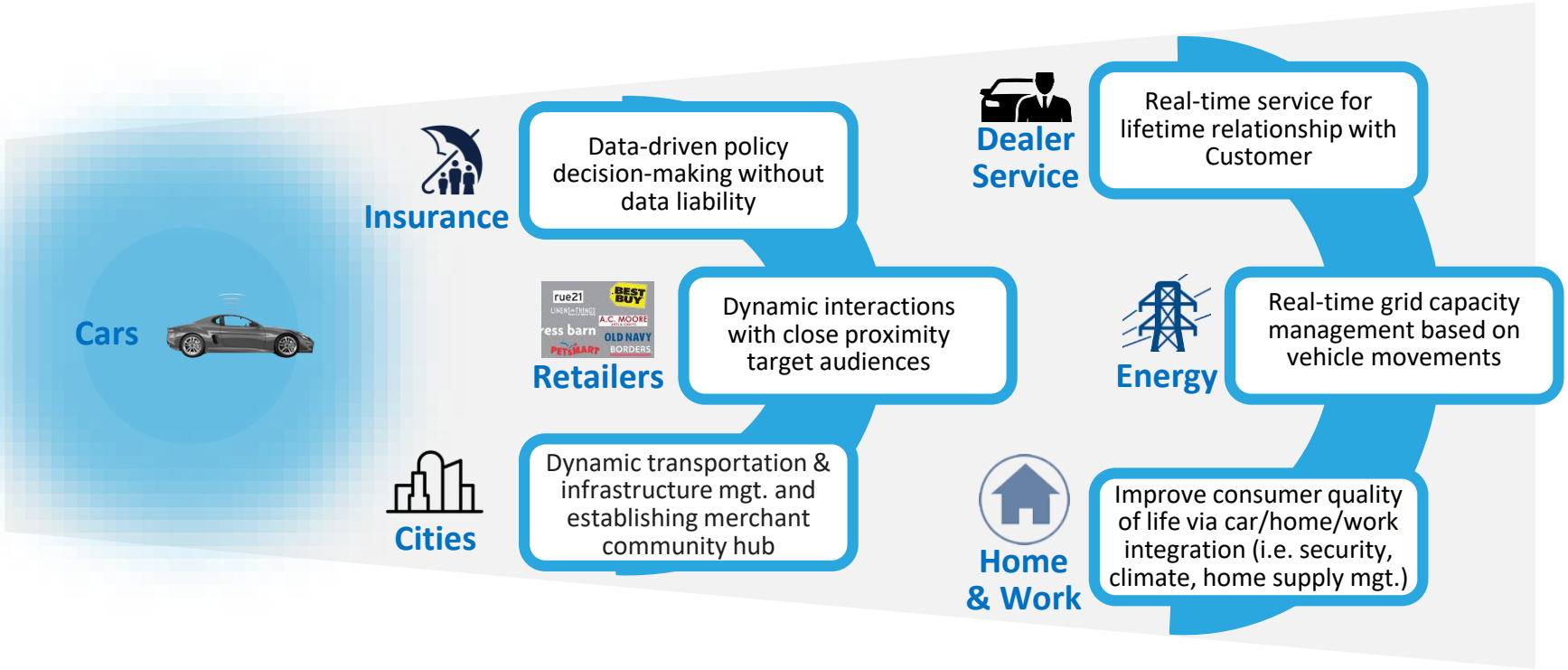
EDGE TO AI

CONNECTED COMMUNITIES

© 2019 Cloudera, Inc. All rights reserved.

CONNECTED VEHICLE IS LEADING THE TRANSFORMATION

Monetizing the Connected Vehicle



© 2019 Cloudera, Inc. All rights reserved.

ESTABLISHING TRUST – WHAT'S REQUIRED?

Data Providers

- Control who can see what data
- Control what data can be combined

Connected Communities

End-Users

ASK QUESTION

DELIVER INSIGHT

Secured views of data and insights

✓ **Data access across systems**

- Each data provider shares specific *data assets* and defines specific data sharing rules

✓ **Controlling who sees what data**

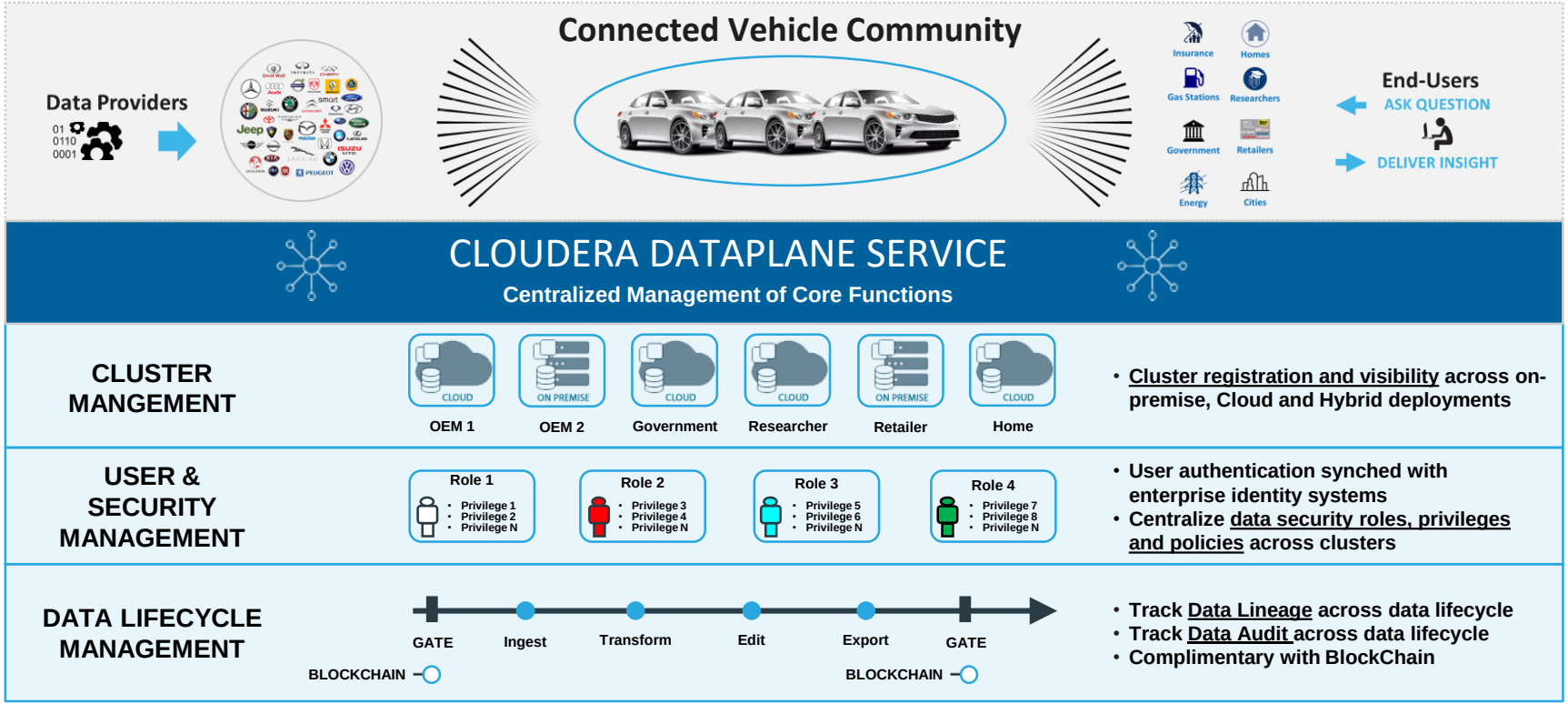
- Each data provider precisely defines *which end-users can see their data* throughout the data lifecycle

✓ **Data audit across the ecosystem**

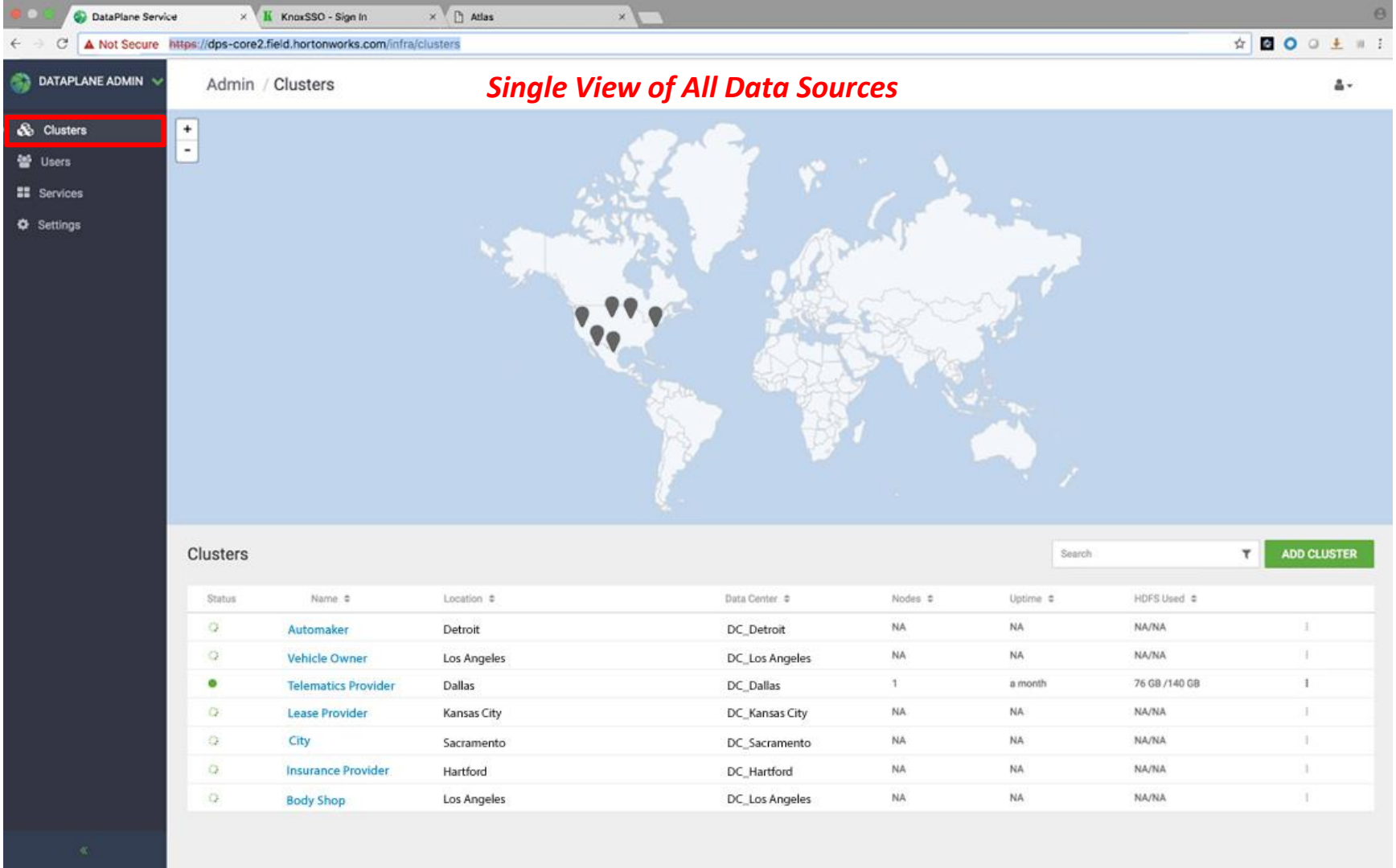
- Each data provider can track *data usage* across the data lifecycle, including the *data's origin, where it is used, and how it changes over time*

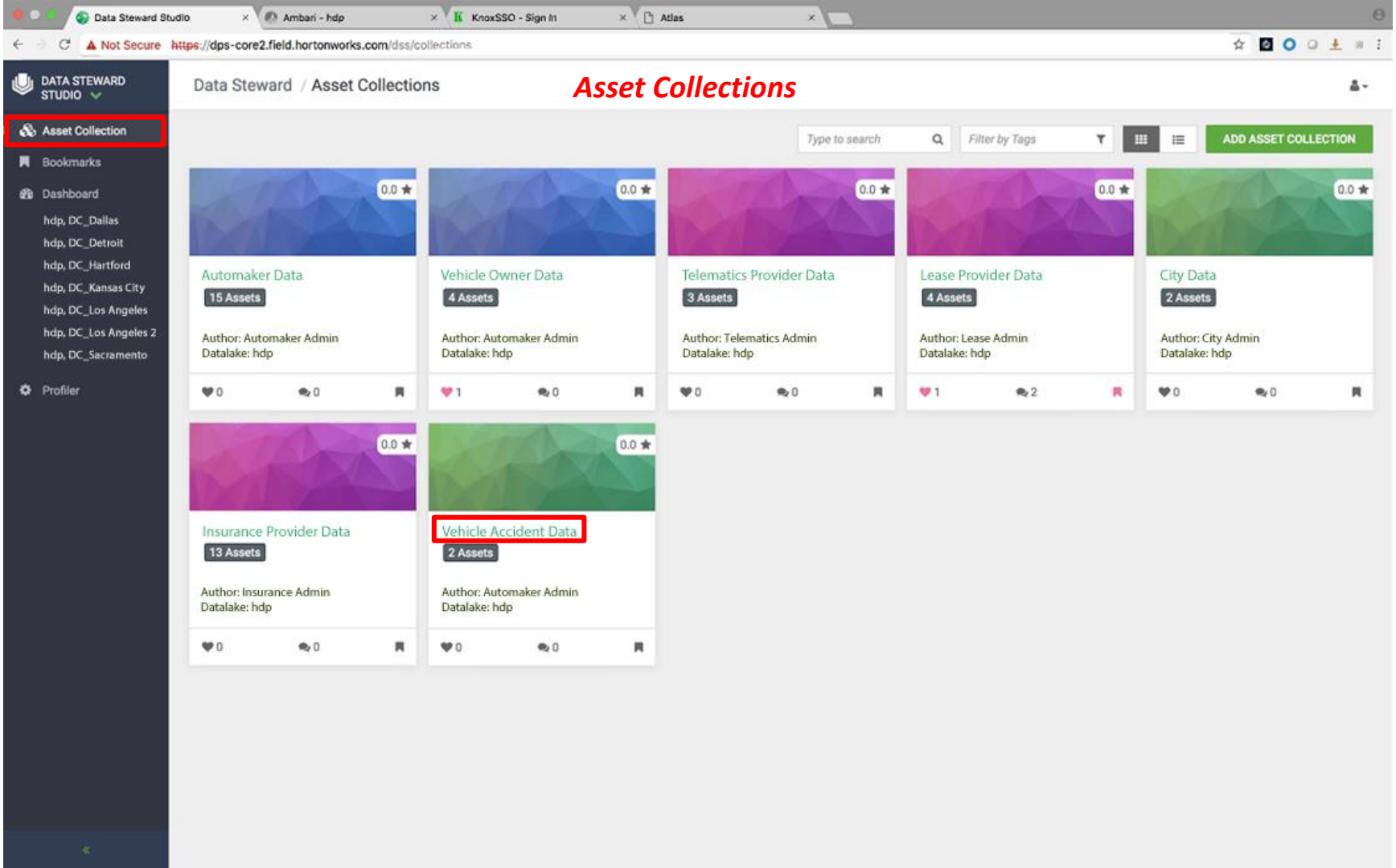
© 2019 Cloudera, Inc. All rights reserved.

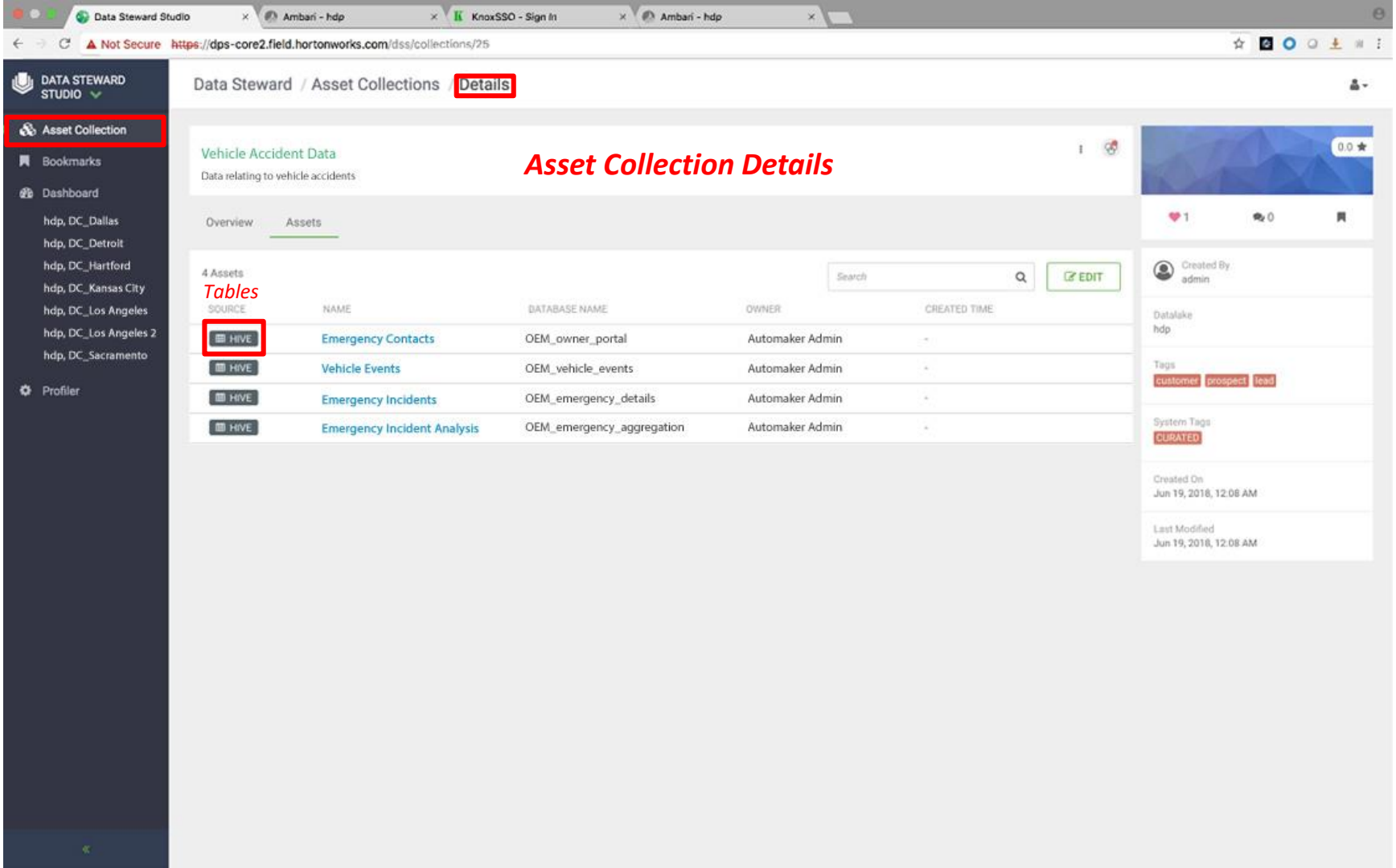
HOW IT WORKS.....

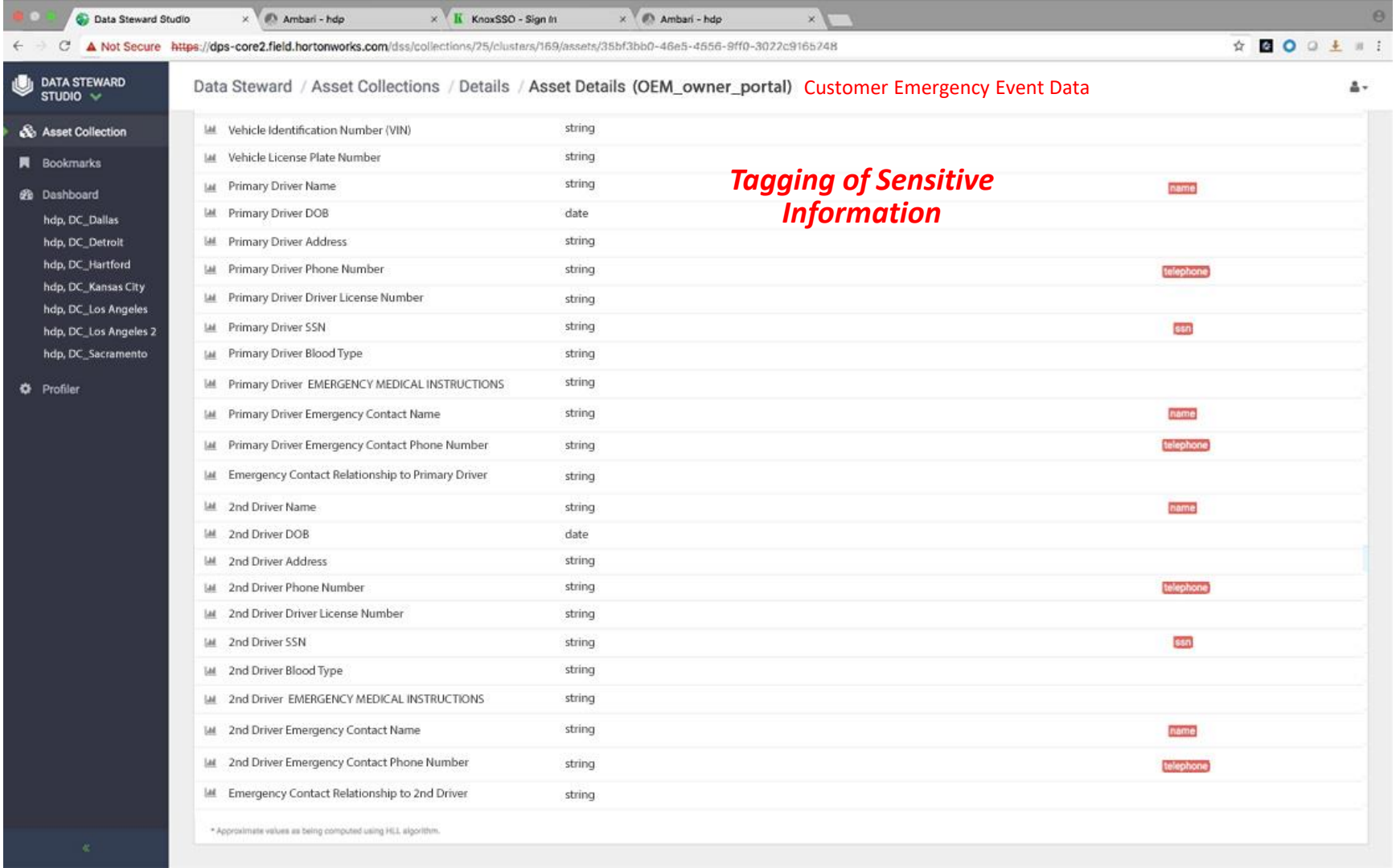


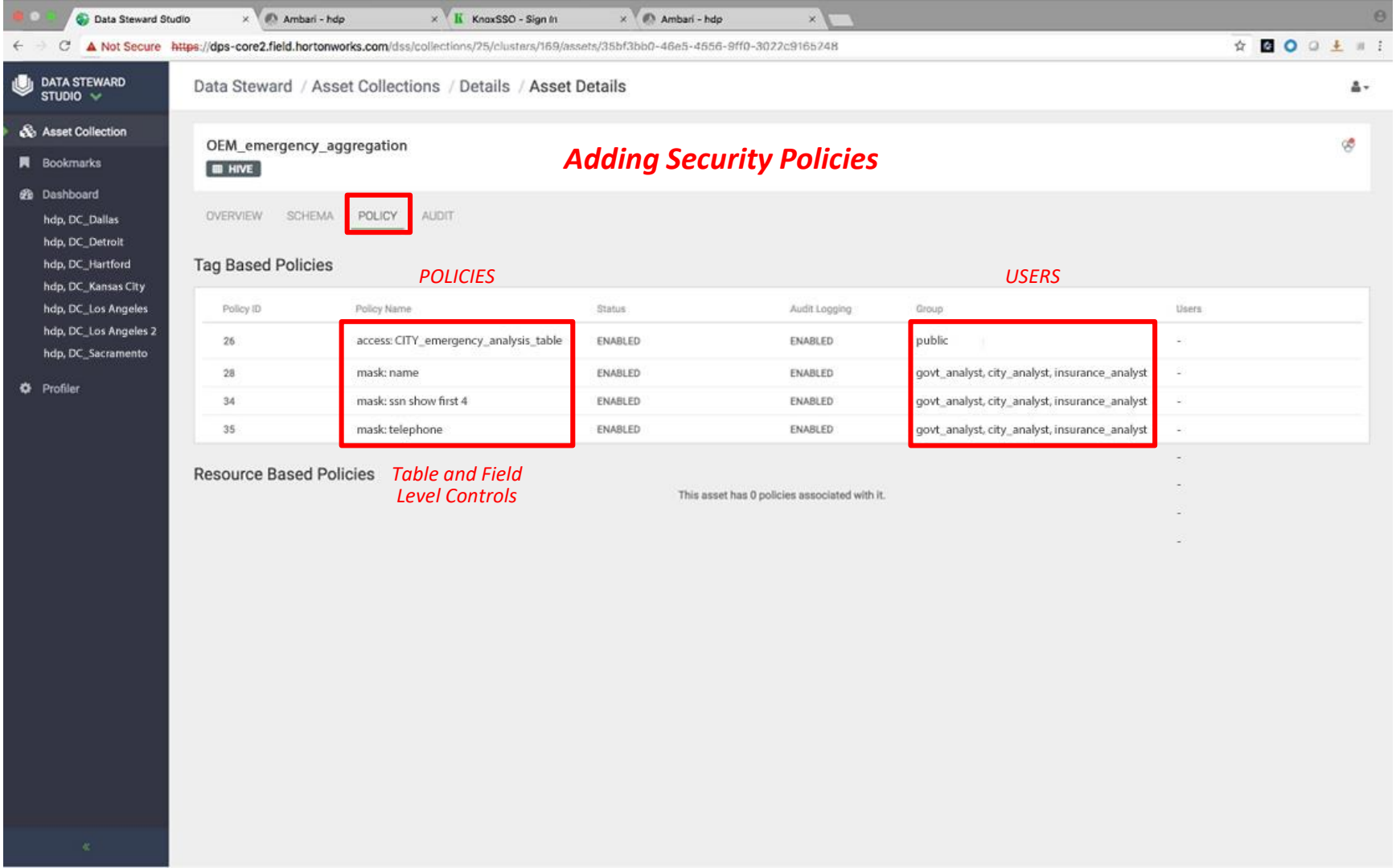
© 2019 Cloudera, Inc. All rights reserved.

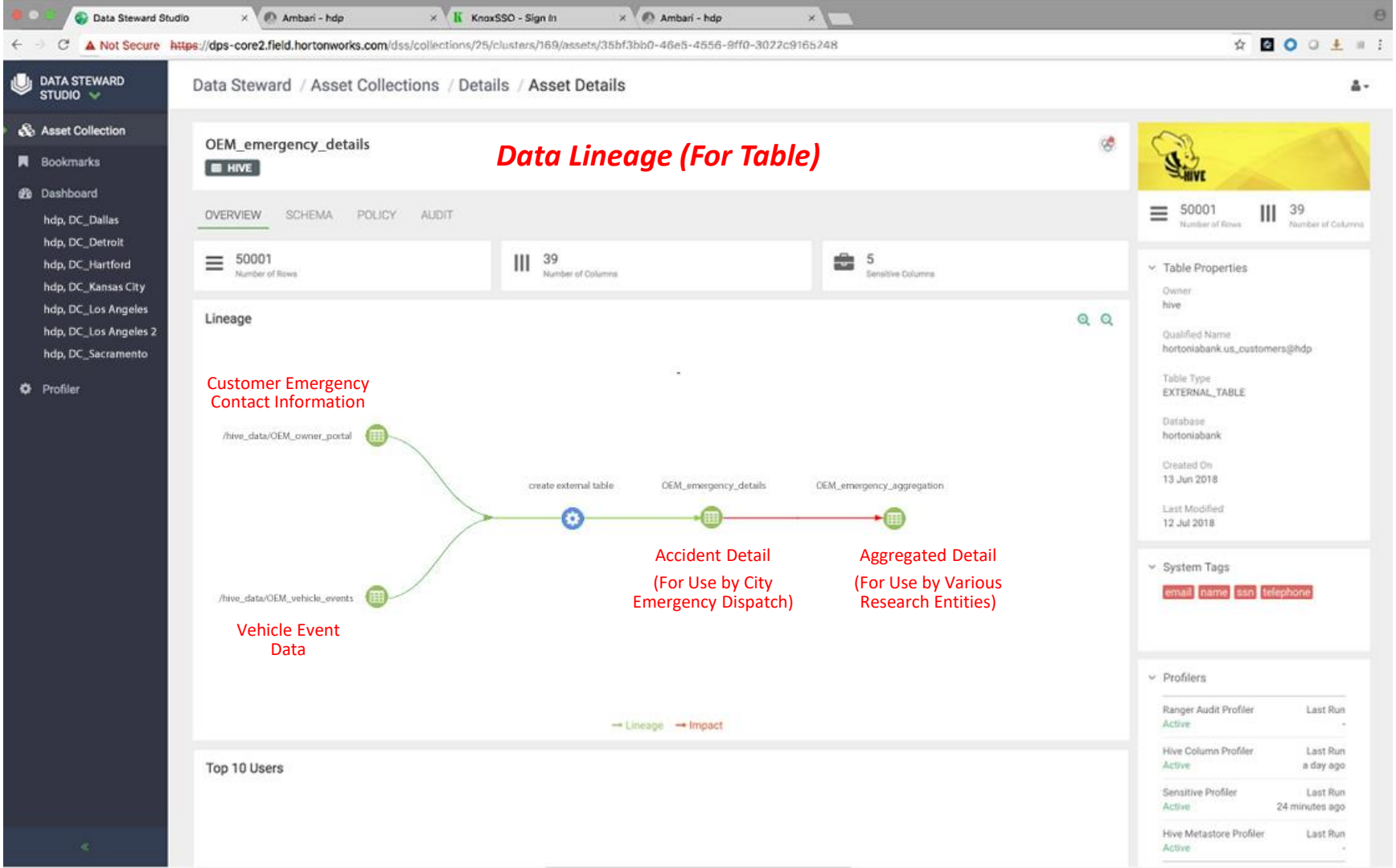

DataPlane Service
KnoxSSO - Sign In
Atlas
Not Secure
https://dps-core2.field.hortonworks.com/infra/clusters
DATAPLANE ADMIN
Clusters
Users
Services
Settings
Admin / Clusters
Single View of All Data Sources
Clusters
Search
ADD CLUSTER
Status
Name
Location
Data Center
Nodes
Uptime
HDFS Used
Automaker
Detroit
DC_Detroit
NA
NA
NA/NA
Vehicle Owner
Los Angeles
DC_Los Angeles
NA
NA
NA/NA
Telematics Provider
Dallas
DC_Dallas
1
a month
76 GB /140 GB
Lease Provider
Kansas City
DC_Kansas City
NA
NA
NA/NA
City
Sacramento
DC_Sacramento
NA
NA
NA/NA
Insurance Provider
Hartford
DC_Hartford
NA
NA
NA/NA
Body Shop
Los Angeles
DC_Los Angeles
NA
NA
NA/NA

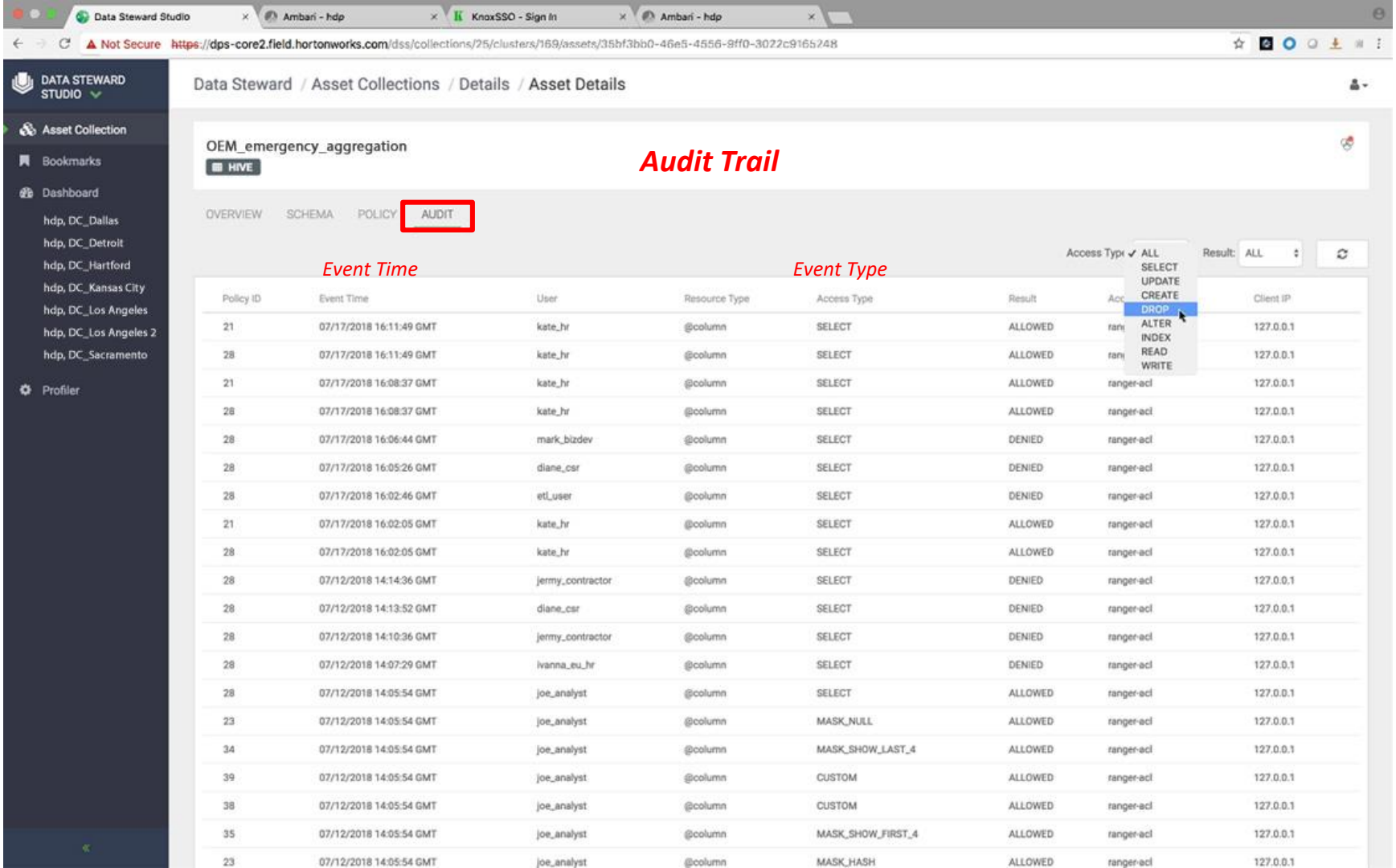

Data Steward Studio
Ambari - hdp
KnoxSSO - Sign In
Atlas
Not Secure https://dps-core2.field.hortonworks.com/dss/collections
DATA STEWARD STUDIO
Asset Collection
Bookmarks
Dashboard
hdp, DC_Dallas
hdp, DC_Detroit
hdp, DC_Hartford
hdp, DC_Kansas City
hdp, DC_Los Angeles
hdp, DC_Los Angeles 2
hdp, DC_Sacramento
Profiler
Data Steward / Asset Collections
Asset Collections
Type to search
Filter by Tags
ADD ASSET COLLECTION
Automaker Data
15 Assets
Author: Automaker Admin
Datalake: hdp
Vehicle Owner Data
4 Assets
Author: Automaker Admin
Datalake: hdp
Telematics Provider Data
3 Assets
Author: Telematics Admin
Datalake: hdp
Lease Provider Data
4 Assets
Author: Lease Admin
Datalake: hdp
City Data
2 Assets
Author: City Admin
Datalake: hdp
Insurance Provider Data
13 Assets
Author: Insurance Admin
Datalake: hdp
Vehicle Accident Data
2 Assets
Author: Automaker Admin
Datalake: hdp


Data Steward Studio
Ambari - hdp
KnoxSSO - Sign In
Ambari - hdp
Not Secure https://dps-core2.field.hortonworks.com/dss/collections/25
DATA STEWARD STUDIO
Asset Collection
Bookmarks
Dashboard
hdp, DC_Dallas
hdp, DC_Detroit
hdp, DC_Hartford
hdp, DC_Kansas City
hdp, DC_Los Angeles
hdp, DC_Los Angeles 2
hdp, DC_Sacramento
Profiler
Data Steward / Asset Collections / Details
Vehicle Accident Data
Data relating to vehicle accidents
Asset Collection Details
Overview
Assets
4 Assets
Tables
Search
EDIT
SOURCE
NAME
DATABASE NAME
OWNER
CREATED TIME
HIVE
Emergency Contacts
OEM_owner_portal
Automaker Admin
HIVE
Vehicle Events
OEM_vehicle_events
Automaker Admin
HIVE
Emergency Incidents
OEM_emergency_details
Automaker Admin
HIVE
Emergency Incident Analysis
OEM_emergency_aggregation
Automaker Admin
0.0
1
0
Created By
admin
Datalake
hdp
Tags
customer
prospect
lead
System Tags
CURATED
Created On
Jun 19, 2018, 12:08 AM
Last Modified
Jun 19, 2018, 12:08 AM


Data Steward Studio
Ambari - hdp
KnoxSSO - Sign In
Ambari - hdp
Not Secure https://dps-core2.field.hortonworks.com/dss/collections/25/clusters/169/assets/35bf3bb0-46e5-4556-9ff0-3022c9165248
DATA STEWARD STUDIO
Asset Collection
Bookmarks
Dashboard
hdp, DC_Dallas
hdp, DC_Detroit
hdp, DC_Hartford
hdp, DC_Kansas City
hdp, DC_Los Angeles
hdp, DC_Los Angeles 2
hdp, DC_Sacramento
Profiler
Data Steward / Asset Collections / Details / Asset Details (OEM_owner_portal) Customer Emergency Event Data
Tagging of Sensitive Information
Vehicle Identification Number (VIN) string
Vehicle License Plate Number string
Primary Driver Name string name
Primary Driver DOB date
Primary Driver Address string
Primary Driver Phone Number string telephone
Primary Driver Driver License Number string
Primary Driver SSN string ssn
Primary Driver Blood Type string
Primary Driver EMERGENCY MEDICAL INSTRUCTIONS string
Primary Driver Emergency Contact Name string name
Primary Driver Emergency Contact Phone Number string telephone
Emergency Contact Relationship to Primary Driver string
2nd Driver Name string name
2nd Driver DOB date
2nd Driver Address string
2nd Driver Phone Number string telephone
2nd Driver Driver License Number string
2nd Driver SSN string ssn
2nd Driver Blood Type string
2nd Driver EMERGENCY MEDICAL INSTRUCTIONS string
2nd Driver Emergency Contact Name string name
2nd Driver Emergency Contact Phone Number string telephone
Emergency Contact Relationship to 2nd Driver string
* Approximate values as being computed using HLL algorithm.


Data Steward Studio
Ambari - hdp
KnoxSSO - Sign In
Ambari - hdp
Not Secure https://dps-core2.field.hortonworks.com/dss/collections/25/clusters/169/assets/35bf3bb0-46e5-4656-9ff0-3022c9165248
DATA STEWARD STUDIO
Asset Collection
Bookmarks
Dashboard
hdp, DC_Dallas
hdp, DC_Detroit
hdp, DC_Hartford
hdp, DC_Kansas City
hdp, DC_Los Angeles
hdp, DC_Los Angeles 2
hdp, DC_Sacramento
Profiler
Data Steward / Asset Collections / Details / Asset Details
OEM_emergency_aggregation
HIVE
Adding Security Policies
OVERVIEW
SCHEMA
POLICY
AUDIT
Tag Based Policies
POLICIES
USERS
Policy ID
Policy Name
Status
Audit Logging
Group
Users
26
access: CITY_emergency_analysis_table
ENABLED
ENABLED
public
28
mask: name
ENABLED
ENABLED
govt_analyst, city_analyst, insurance_analyst
34
mask: ssn show first 4
ENABLED
ENABLED
govt_analyst, city_analyst, insurance_analyst
35
mask: telephone
ENABLED
ENABLED
govt_analyst, city_analyst, insurance_analyst
Resource Based Policies
Table and Field Level Controls
This asset has 0 policies associated with it.


Data Steward Studio
Ambari - hdp
KnoxSSO - Sign In
Ambari - hdp
Not Secure https://dps-core2.field.hortonworks.com/dss/collections/25/clusters/169/assets/35bf3bb0-46e5-4556-9ff0-3022c9165248
DATA STEWARD STUDIO
Asset Collection
Bookmarks
Dashboard
hdp, DC_Dallas
hdp, DC_Detroit
hdp, DC_Hartford
hdp, DC_Kansas City
hdp, DC_Los Angeles
hdp, DC_Los Angeles 2
hdp, DC_Sacramento
Profiler
Data Steward / Asset Collections / Details / Asset Details
OEM_emergency_details
HIVE
Data Lineage (For Table)
OVERVIEW
SCHEMA
POLICY
AUDIT
50001
Number of Rows
39
Number of Columns
5
Sensitive Columns
Lineage
Customer Emergency Contact Information
/hive_data/OEM_owner_portal
/hive_data/OEM_vehicle_events
Vehicle Event Data
create external table
OEM_emergency_details
OEM_emergency_aggregation
Accident Detail
(For Use by City Emergency Dispatch)
Aggregated Detail
(For Use by Various Research Entities)
Lineage
Impact
Top 10 Users
50001
Number of Rows
39
Number of Columns
Table Properties
Owner
hive
Qualified Name
hortoniabank.us_customers@hdp
Table Type
EXTERNAL_TABLE
Database
hortoniabank
Created On
13 Jun 2018
Last Modified
12 Jul 2018
System Tags
email
name
ssn
telephone
Profilers
Ranger Audit Profiler
Active
Last Run
Hive Column Profiler
Active
Last Run
a day ago
Sensitive Profiler
Active
Last Run
24 minutes ago
Hive Metastore Profiler
Active
Last Run


Data Steward Studio
Ambari - hdp
KnoxSSO - Sign In
Ambari - hdp
Not Secure https://dps-core2.field.hortonworks.com/dss/collections/25/clusters/169/assets/35bf3bb0-46e5-4556-9ff0-3022c9165248
DATA STEWARD STUDIO
Asset Collection
Bookmarks
Dashboard
hdp, DC_Dallas
hdp, DC_Detroit
hdp, DC_Hartford
hdp, DC_Kansas City
hdp, DC_Los Angeles
hdp, DC_Los Angeles 2
hdp, DC_Sacramento
Profiler
Data Steward / Asset Collections / Details / Asset Details
OEM_emergency_aggregation
HIVE
Audit Trail
OVERVIEW
SCHEMA
POLICY
AUDIT
Event Time
Event Type
Access Type: ALL
SELECT
UPDATE
CREATE
DROP
ALTER
INDEX
READ
WRITE
Result: ALL
Policy ID | Event Time | User | Resource Type | Access Type | Result | Client IP
21 | 07/17/2018 16:11:49 GMT | kate_hr | @column | SELECT | ALLOWED | 127.0.0.1
28 | 07/17/2018 16:11:49 GMT | kate_hr | @column | SELECT | ALLOWED | 127.0.0.1
21 | 07/17/2018 16:08:37 GMT | kate_hr | @column | SELECT | ALLOWED | ranger-acl | 127.0.0.1
28 | 07/17/2018 16:08:37 GMT | kate_hr | @column | SELECT | ALLOWED | ranger-acl | 127.0.0.1
28 | 07/17/2018 16:06:44 GMT | mark_bizdev | @column | SELECT | DENIED | ranger-acl | 127.0.0.1
28 | 07/17/2018 16:05:26 GMT | diane_csr | @column | SELECT | DENIED | ranger-acl | 127.0.0.1
28 | 07/17/2018 16:02:46 GMT | etl_user | @column | SELECT | DENIED | ranger-acl | 127.0.0.1
21 | 07/17/2018 16:02:05 GMT | kate_hr | @column | SELECT | ALLOWED | ranger-acl | 127.0.0.1
28 | 07/17/2018 16:02:05 GMT | kate_hr | @column | SELECT | ALLOWED | ranger-acl | 127.0.0.1
28 | 07/12/2018 14:14:36 GMT | jermy_contractor | @column | SELECT | DENIED | ranger-acl | 127.0.0.1
28 | 07/12/2018 14:13:52 GMT | diane_csr | @column | SELECT | DENIED | ranger-acl | 127.0.0.1
28 | 07/12/2018 14:10:36 GMT | jermy_contractor | @column | SELECT | DENIED | ranger-acl | 127.0.0.1
28 | 07/12/2018 14:07:29 GMT | ivanna_eu_hr | @column | SELECT | DENIED | ranger-acl | 127.0.0.1
28 | 07/12/2018 14:05:54 GMT | joe_analyst | @column | SELECT | ALLOWED | ranger-acl | 127.0.0.1
23 | 07/12/2018 14:05:54 GMT | joe_analyst | @column | MASK_NULL | ALLOWED | ranger-acl | 127.0.0.1
34 | 07/12/2018 14:05:54 GMT | joe_analyst | @column | MASK_SHOW_LAST_4 | ALLOWED | ranger-acl | 127.0.0.1
39 | 07/12/2018 14:05:54 GMT | joe_analyst | @column | CUSTOM | ALLOWED | ranger-acl | 127.0.0.1
38 | 07/12/2018 14:05:54 GMT | joe_analyst | @column | CUSTOM | ALLOWED | ranger-acl | 127.0.0.1
35 | 07/12/2018 14:05:54 GMT | joe_analyst | @column | MASK_SHOW_FIRST_4 | ALLOWED | ranger-acl | 127.0.0.1
23 | 07/12/2018 14:05:54 GMT | joe_analyst | @column | MASK_HASH | ALLOWED | ranger-acl | 127.0.0.1

CONCLUSION

- Connected Vehicle Data is Key to Industry Transformation
- Evolving Intelligence at Edge and Connected Community Use Cases
- Big Data Management Challenges on Multiple Levels
- Open Source Data Management Innovation Can Help!

© 2019 Cloudera, Inc. All rights reserved.

THANK YOU

CLOUDERA