SECURING DATA
WITH DATAGUISE
DgSecure

DATAGUISE

Use Cases

Table of Contents

**Introduction: Using DgSecure in the Hortonworks Sandbox**

This document will guide you through several use cases around Discovering and protecting sensitive data.   Discovery / Masking / Encryption & Decryption.   The hadoop cluster is hidden behind the scenes during this evaluation.

Getting Started

When the interface shows '**TRIAL IN PROGRESS**", you should also receive an email from Dataguise labeled: 'Dataguise test drive Trial Details' - informing you that the 5 hour trial period has started. Do not wait for the 'TRIAL COMPLETE ' entry to start testing – that means the test period is over.

Trial In-progress
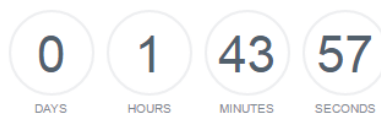
Remaining

0	1	43	57
DAYS	HOURS	MINUTES	SECONDS

Your DgSecure for HDInsight test drive is ready! Follow the instructions in the user guide and use the credentials sent to your email and the environment window to get started.

Stop

**The hour, minutes, and seconds evaluations show your remaining time.**

In that email will be two things, a link to the test drive User Guide, and a starting URL link for DgSecure.

Click on this link to bring you to the DgSecure login Screen

DATAGUISE®
DgSecure

Username: dataguise
Password: ●●●●●●●●

Sign In

Forgot Password?

Logging into DgSecure interface

- o   Enter username: **dataguise**
- o   Enter password: **dataguise**
- o   You will get a message about the limited license:
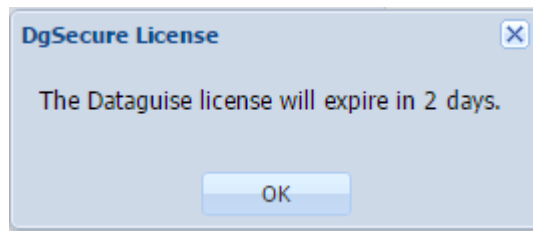
This is actually a 5 hour license but the 2 days covers 5 hour trials that span over midnight.
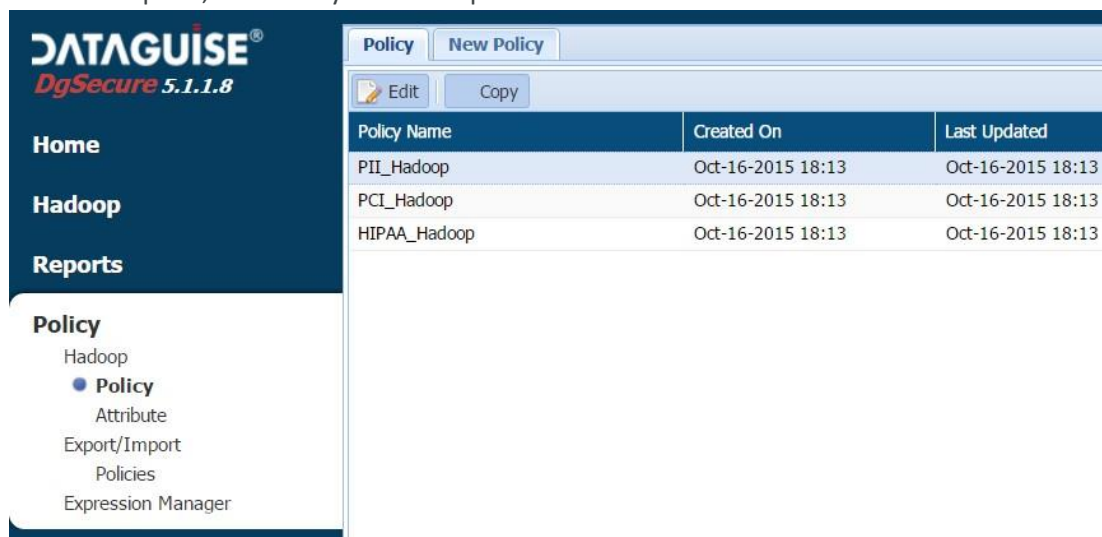
o On the left side of the DgSecure interface, you will see these sections:

We will start with the Policy Section

Create Policy

o On the left pane, click Policy – it will expand and the screen should be similar to below

o On the top pane, click "New Policy" tab
o Fill in the Policy name: **Tutorial**

- o Fill in the Description field: **Tutorial**
- o Check the following sensitive elements and with each element, choose the masking options for each. The defaults is FPM (Format Preserving Masking).   Choose all Social Security and all Telephone entries only for this test. <u>Do not</u> check any of the consistency flags on the right.

Policy	New Policy		

Policy:	Tutorial		Description:	Tutorial	

|	Expression	Description	Masking Option
☐	Credit Card # (Space Separation)	e.g. 5173-2157-5085-6134	Select Masking Option
☐	Credit Card # (Dash Separation)	e.g. 5173-2157-5085-6134	Select Masking Option
☐	**Names**		
☐	Part Names(FirstName or LastName)	Part Names(First Name,Last Name, First/LastName)	Select Masking Option
☐	Full Names	Full Names	Select Masking Option
☑	**Social Security**		
☑	Social Security # (Digits Only)	e.g. 232883211	FPM
☑	Social Security # (Space Separation)	e.g. 232 88 3211	FPM
☑	Social Security # (Dash Separation)	e.g. 232-88-3211	FPM
☑	**Telephone**		
☑	Telephone (Digits Only)	e.g. 8776320522	FPM
☑	Telephone (Space Separation)	e.g. 877 632 0522	FPM
☑	Telephone (Dash Separation)	e.g. 877-632-0522	FPM
☑	Telephone (Standard)	e.g. (877) 632-0522	FPM
☑	Telephone (Standard without spaces)	e.g. (877)632-0522	FPM

- o Click Save (bottom right)
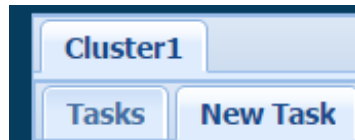- o Policy is now created and ready to use

Use Cases

1. **Unstructured Discovery**

   Setup & Run Discovery task
   - o Click on Hadoop tab in the left pane. This will highlight the "Tasks" tab

DATAGUISE
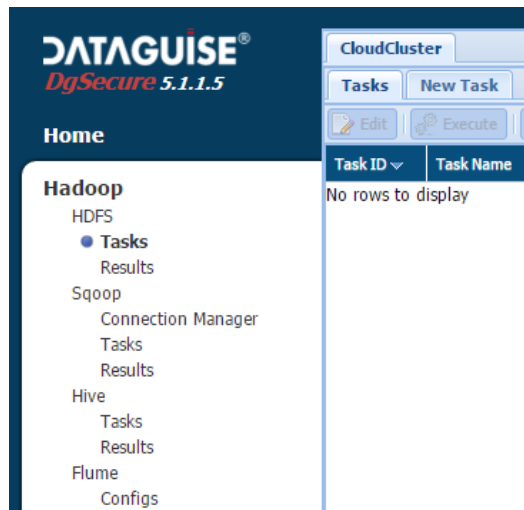
- On the right window click "New Task"

- Fill in the following fields as shown in the figure below

  - Task Name: **UnstructDiscovery**

  - Task Description**: UnstructDiscovery**

  - Check the 'Full Scan' box

  - Task Type: **Discovery only** (select from the dropdown)

  - Select a Compliance Policy. We will be using "**Tutorial**". When you check the Tutorial Policy, the entries you saved (SSN and Telephone) should be listed in blue below.

  - On the right of this new task windows, click on the "Select Directories" button.

- Select the directories you would like to discover sensitive data in and click **Add**. We will be using "/HR". Click on HR and press Add.

- The HR directory will now be listed in the bottom window.

- After adding all desired directories, click **Done**. This will bring back to the new task window
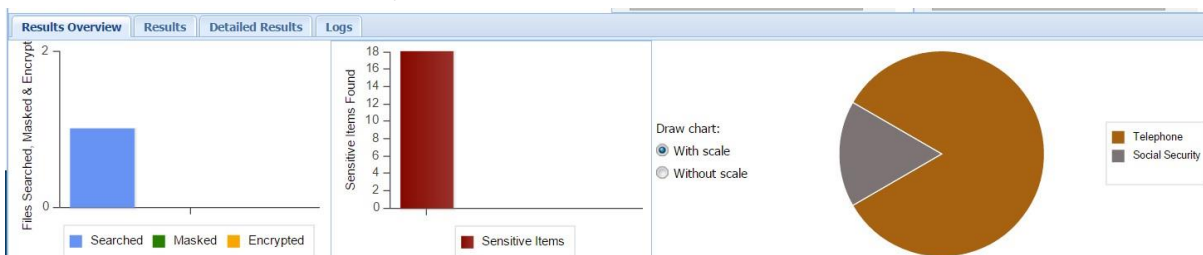
- Click "Save and Execute" button in the bottom right corner of this window
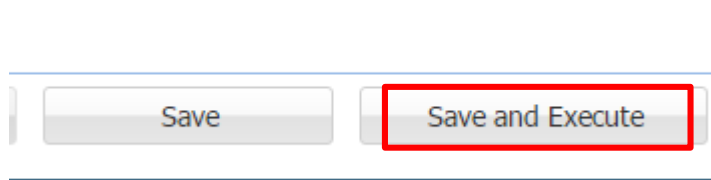
- The task will be submitted you will be taken to the "Results" window
- The task will go through several phases, like "Initializing", "Running" and "Summarizing Results"
- The Task will be started (actually submits a MapReduce job)
- This task should take about 1 minute.

- Once task is finished, it will show results as follows.

- You can hover over the Pie Chart to see counts of each. Your policy only checked for SSN and telephone numbers in the Discovery – so only those two sensitive attributes should be listed.
- A detailed result can also be viewed by clicking on the "Detailed Results" tab.

- You can also get (CSV, XLS) as well as PDF reports by clicking on the "Save Results to File" or "Save Results to PDF" buttons in top right corner of the bottom pane

- Excel Spreadsheet

	A	B	C	D	E	F	G	H	I	J
1	Task Instance ID	Task Name	Sensitive Type	File Path	Fil	Di	C	Sensitive Items Count	Masked	Encrypted
2	1	UnstructDiscovery	Telephone (Space Separation)	//cl29104@st1cl622i29104.blob.core.windows.net/HR/charter_one.txt				3	FALSE	FALSE
3	1	UnstructDiscovery	Social Security # (Space Separation)	//cl29104@st1cl622i29104.blob.core.windows.net/HR/charter_one.txt				1	FALSE	FALSE
4	1	UnstructDiscovery	Telephone (Digits Only)	//cl29104@st1cl622i29104.blob.core.windows.net/HR/charter_one.txt				3	FALSE	FALSE
5	1	UnstructDiscovery	Telephone (Standard without spaces)	//cl29104@st1cl622i29104.blob.core.windows.net/HR/charter_one.txt				3	FALSE	FALSE
6	1	UnstructDiscovery	Social Security # (Digits Only)	//cl29104@st1cl622i29104.blob.core.windows.net/HR/charter_one.txt				1	FALSE	FALSE
7	1	UnstructDiscovery	Telephone (Dash Separation)	//cl29104@st1cl622i29104.blob.core.windows.net/HR/charter_one.txt				3	FALSE	FALSE
8	1	UnstructDiscovery	Social Security # (Dash Separation)	//cl29104@st1cl622i29104.blob.core.windows.net/HR/charter_one.txt				1	FALSE	FALSE
9	1	UnstructDiscovery	Telephone (Standard)	//cl29104@st1cl622i29104.blob.core.windows.net/HR/charter_one.txt				3	FALSE	FALSE
10										

- PDF Report

HDFS Task Results
Generated By : admin On : May-20-2015 08:46

Task Details

Task Name : Discover
Task Type : Discovery
Directory Path(s) Selected :
/Engineering
/Marketing
Sensitive Data Type(s) Selected :

Task Instance Id : 1
Task Description : Discover
Policies Selected : PII_Hadoop

Structured : false

Policy Names	Policy Expressions
PII_Hadoop	[Address (Unstructured data only), Social Security # (Digits Only), Social Security # (Space Separation), Social Security # (Dash Separation), Telephone (Digits Only), Telephone (Space Separation), Telephone (Dash Separation), Telephone (Standard), Telephone (Standard without spaces), Names, Email Address]

Detailed Results

Sensitive Data Type	File Path	Sensitive Items Count	Masked	Encrypted	Incremental Scan
Telephone (Space Separation)	/Engineering/StructuredData-I.txt	7	No	No	false
Social Security # (Dash Separation)	/Engineering/StructuredData-I.txt	9	No	No	false

Unstructured Discovery with more sensitive elements

- You can rerun this same task (by editing it) but check the PII_hadoop Policy. This has many more sensitive elements in the Policy window by default (lower left below)

DATAGUISE

Compliance Policies

☐ HIPAA_Hadoop   ☐ PCI_Hadoop   ☑ PII_Hadoop   ☐ Tutorial

Pre-defined and Custom Expressions		Manage Scan locations

Pre-defined and Custom Expressions

☐ Name	Description
Full Names	Full Names
☐ Part Names(FirstName or LastName)	Part Names(First Name,Last Name, First/LastName)
☐ ☑ **Social Security**	
☐ Social Security # (Digits Only)	e.g. 232883211
☐ Social Security # (Space Separation)	e.g. 232 88 3211
☐ Social Security # (Dash Separation)	e.g. 232-88-3211
☐ ☑ **Telephone**	
☐ Telephone (Digits Only)	e.g. 8776320522
☐ Telephone (Space Separation)	e.g. 877 632 0522
☐ Telephone (Dash Separation)	e.g. 877-632-0522
☐ Telephone (Standard)	e.g. (877) 632-0522
☐ Telephone (Standard without spaces)	e.g. (877)632-0522
☐ ABA Routing number	Routing number for ABA
☐ Email Address	Email Address
☐ IP Address	IP Address
☐ NPI	NPI
☐ URL	URL

Manage Scan locations

[Select Directories]

Scan Location

/HR

[Cancel]   [Save As]

- ■ The results show many more discovered sensitive elements

- ■ Names, Addresses and Email Addresses – in addition to SSN and Telephone

Legend:
- Email Address
- Names
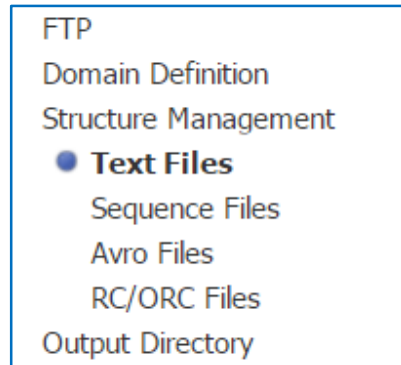- Telephone
- Address
- Social Security
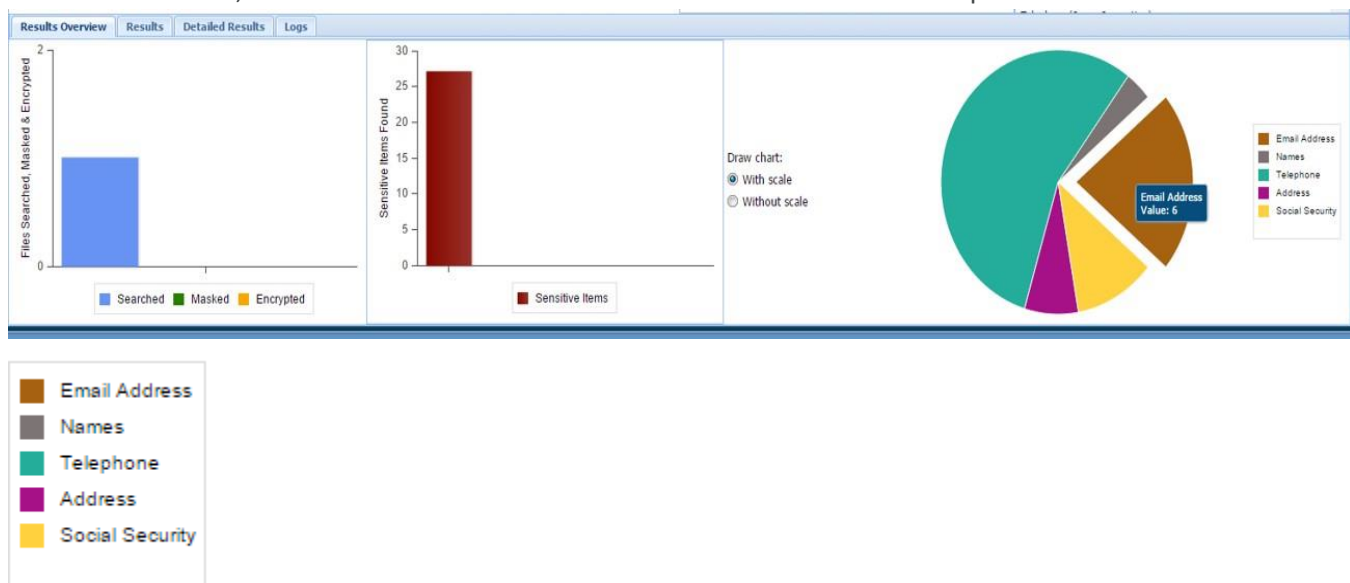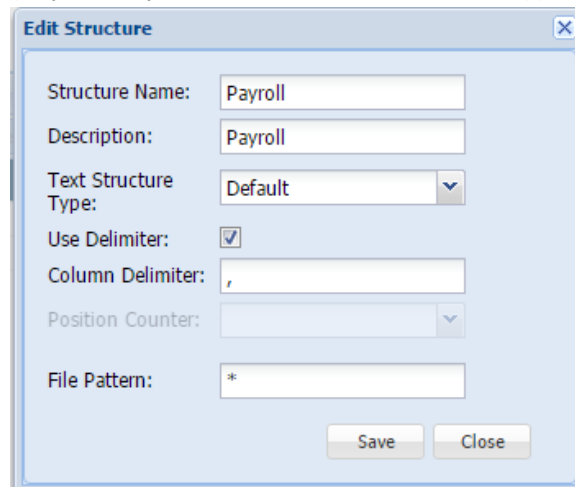
2. Structured Discovery

Setup Structure

Defining structures tells DgSecure exactly where the sensitive elements are located.

- Click on "Text File" under Structure Management tab in the left pane

FTP
Domain Definition
Structure Management
● **Text Files**
Sequence Files
Avro Files
RC/ORC Files
Output Directory

-
- Click "Add Structure" tab
- Fill out the form like the following

    (Payroll, Payroll, Column Delimiter comma (,), File Pattern asterisk (*))

Edit Structure

Structure Name:	Payroll
Description:	Payroll
Text Structure Type:	Default
Use Delimiter:	☑
Column Delimiter:	,
Position Counter:	
File Pattern:	*

Save   Close

- Click **Save**
- Next setup Columns for the structure by clicking "**Add Column**" tab in the middle pane
- Fill the form as follows
- SSN. Column 1 and Social Security # (Digits Only)

- Click **Save**

- Next click add column again to add another column, fill out the form as follows.
- TEL, Column 8 and Telephone (Digits Only)

- Click **Save**

- You should have these 2 columns in your structure

Column Number	Column Name	Expression Type
1	SSN	Social Security # (Digits Only)
8	TEL	Telephone (Digits Only)

-

- Now associate a directory to the structure by clicking "Add Directory" and filling out the form as follows. **Browse to select the /Marketing Dir**. Pulldown Structure → Payroll

DATAGUISE
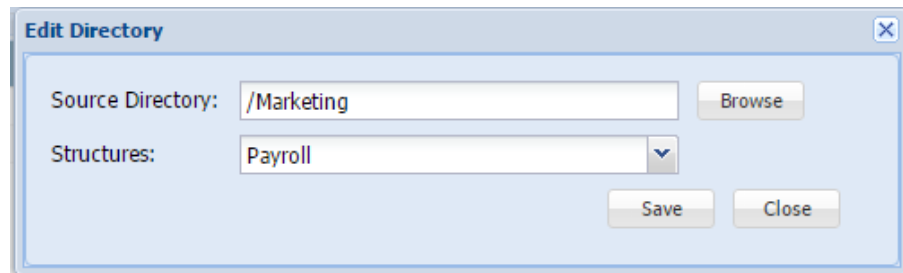
- Click **Save**
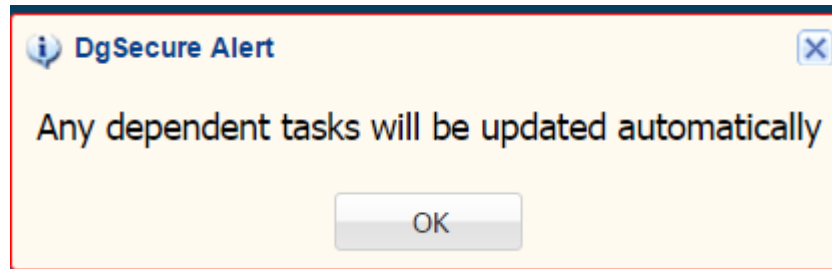
- Last step is to **Save the structure** by clicking the "**Save Structure**" button on the top pane, (which should be flashing red) as shown below

-
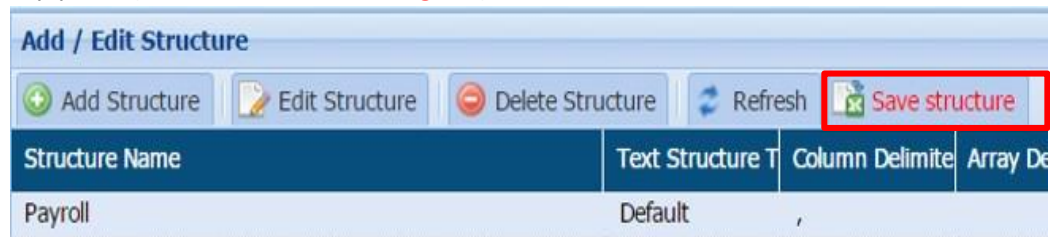
Run Structured Discovery
- Create a new task as previously created →**Hadoop → Tasks → New Task**
- When selecting the directory, choose the one that has a structure associated with it.
  In this case we'll use "**/Marketing**"
  - Task Name: **StructDiscovery**
  - Task Description: **StructDiscovery**
  - Task Type: **Discovery** (from the drop down)
  - Full Scan: **Check the box**
  - Compliance Policy: **Tutorial**
  - Select Directories: **Marketing**
  - **Save and Execute**

- **Save and Execute** the task

Review Results

Results for Structured Discovery can be seen in the results tab. The results show the number of files searched (BLUE), total number of sensitive elements found (RED) and a pie chart with breakdown of the types of sensitive elements found.

- Detailed results can be seen under "Detailed Results" tab

| | | | | | | Clear Filters | Save Results to File | Save Results to Pd |

Task Name	Sensitive Data Type ▲	File Path	File Typ	Discover	Content	Hit Count	Masked	Encrypted	Incremental Sc
StructDiscovery	Social Security # (Digits Only)	//cl29104@st1cl622i29104.blob.core.windows.net/Marketing/payroll.csv				100	false	false	false
StructDiscovery	Telephone (Digits Only)	//cl29104@st1cl622i29104.blob.core.windows.net/Marketing/payroll.csv				40	false	false	false

- A PDF or consumable report can also be downloaded

HDFS Task Results
Generated By : admin On : Sep-18-2015 17:37

Task Details

Task Name : StructDiscovery
Task Type : Discovery
Directory Path(s) Selected :
  /Marketing
Sensitive Data Type(s) Selected :

Task Instance Id : 32
Task Description : StructDiscovery
Policies Selected : ssn&tel

Structured : false

Policy Names	Policy Expressions
ssn&tel	[Telephone (Digits Only), Telephone (Space Separation), Telephone (Dash Separation), Telephone (Standard), Telephone (Standard without spaces), Social Security # (Digits Only), Social Security # (Space Separation), Social Security # (Dash Separation)]

Detailed Results

Sensitive Data Type	File Path	File Type	Discovery Type	Content Read	Sensitive Items Count	Masked	Encrypted	Increment al Scan
Social Security # (Digits Only)	//dghdicluster@dghdistorage.blob.core.windows.net/Marketing/payroll 2.csv				100	No	No	false
Telephone (Digits Only)	//dghdicluster@dghdistorage.blob.core.windows.net/Marketing/payroll 2.csv				40	No	No	false

3. Unstructured Masking

Setup Output Directory

Before we setup an unstructured masking task, we first need to create an Output Directory. The output directory is the directory where all the masked files will be stored. If you don't setup an output directory, then the files will be created in the default output directory according to the setting in the HDFS configuration file.

- Click on the Hadoop tab in the left pane, then click "**Output Directory**"
- Click on "**Add Source Directory**"
- For the "Source Directory", click on "Browse" button and select the directory that has files that needs to be masked. In this case, select "**Sales**"
- Branch Point is branch off of the destination directory to place output masked file in a specific folder.
- Destination directory is where masked output file will be dropped

DATAGUISE

- ▪ Click **Save**

Run Un-Structured Masking Task

- o Create new Hadoop task (as in previous steps) by click Hadoop -> task -> new task
- o Define the task as follow
    - ▪ Task Name: **UnstructuredMasking**
    - ▪ Task Description: **UnstructuredMasking**
    - ▪ Task Type: **Masking** (from the drop down)
    - ▪ Full Scan: **Check the box**
    - ▪ Structured: DO NOT check the box
    - ▪ Compliance Policy: **Tutorial**
    - ▪ Select Directories: **Sales**
    - ▪ **Save and Execute**

Review Results

- Results for unstructured masking would display as follows. The number of files that were searched (BLUE), the number of files that were masked (GREEN), the number of sensitive data items found in those files (RED), and a pie chart with the breakdown of the different types of sensitive elements.
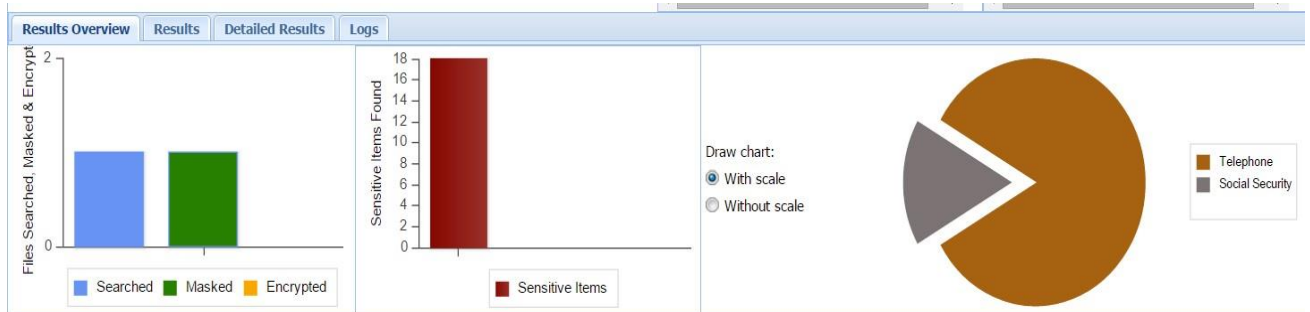
- Click on the "**Detailed Results**" tab on top left of the bottom pane. This will show the details of the scan, which includes Sensitive Data Type, location as well as protection status.

Results Overview	Results	Detailed Results	Logs							
						Clear Filters	Save Results to File	Save Results to		
Task Name	Sensitive Data Type ▲	File Path		File Typ	Discovery	Content R₁	Hit Count	Masked	Encrypted	Incremental
UnstructuredMasking	Social Security # (Dash Separatio...	//cl29104@st1cl622i29104.blob.core.windows.net/Sales/charter_one.txt					1	true	false	false
UnstructuredMasking	Social Security # (Digits Only)	//cl29104@st1cl622i29104.blob.core.windows.net/Sales/charter_one.txt					1	true	false	false
UnstructuredMasking	Social Security # (Space Separati...	//cl29104@st1cl622i29104.blob.core.windows.net/Sales/charter_one.txt					1	true	false	false
UnstructuredMasking	Telephone (Dash Separation)	//cl29104@st1cl622i29104.blob.core.windows.net/Sales/charter_one.txt					3	true	false	false
UnstructuredMasking	Telephone (Digits Only)	//cl29104@st1cl622i29104.blob.core.windows.net/Sales/charter_one.txt					3	true	false	false
UnstructuredMasking	Telephone (Space Separation)	//cl29104@st1cl622i29104.blob.core.windows.net/Sales/charter_one.txt					3	true	false	false

- This is a limited evaluation to start learning of the capabilities of DgSecure. You do not have access to the created directories for this test – however the **/MaskSalesOut** directory was created in Hadoop and the 1 file in that directory (charter_one.txt) was also created but the SSN and Telephone Numbers were masked.

- A sample BEFORE and FILE CREATE AFTER MASKING file image below shows an unstructured text file – charter_one.txt. This file contains numerous sensitive items, names, address, telephone numbers, emails, SSN and a custom element. The image below shows the before and after image of the masked file – in this case – masking all sensitive items – not just SSN and telephone number in our test above.
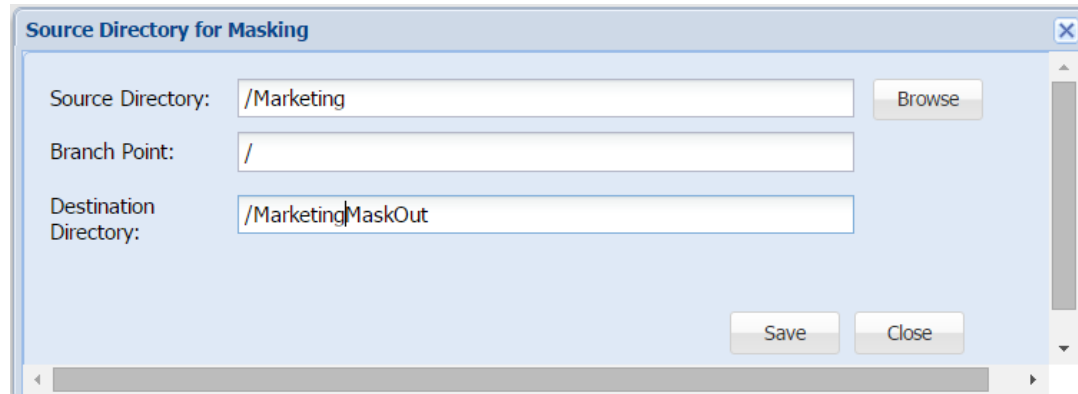
BEFORE   **FILE CREATED AFTER MASKING**

To	:	Charter-Four Auto Finance	To	:	Charter-Four Auto Finance
		228 East Main Street Suite 300 Rochester, NY			95436 Beacher Blvd Suite 367 Harbine Mi 19641
		Attn: Payoff Dept			Attn: Payoff Dept
From	:	Paul Capobianco	From	:	Collice Doerr
		79 Audubon Drive Walpole, MA 02081			13575 Osama Blvd # 561 Albion New York 15800
		Home: (508) 660-3098			Home: (423) 922-8703
		Bus: (508) 850-0058			Bus: (618) 702-7558
		Fax: (508) 660-3098 (When plugged in)			Fax: (620) 436-4555 (When plugged in)
Date	:	Thursday, December 17th, 2015	Date	:	Thursday, December 17th, 2015
Email	:	paul.capobianco@dataguise.com My s	Email	:	ADAMINA@OLIVE.EDU My secondary e
Subj	:	Unknown Transaction	Subj	:	Unknown Transaction

=======================================   =======================================

To Payoff Dept,

    I called and received a payoff figure of $1,272.02,

the purpose of an unknown charge on the account on October 1

not familiar. Could you verify this transaction and gat bac

number which is 013-52-1234 and also under Account# ZX-07672

or emails. I should be available anytime for your response.

To Payoff Dept,

    I called and received a payoff figure of $1,272.02, g

the purpose of an unknown charge on the account on October 12,

not familiar. Could you verify this transaction and gat back

number which is 735-27-8660 and also under Account# FF-20264.

or emails. I should be available anytime for your response.

4. Structured Masking

Run Structured Masking

- o Create output directory
    - Click on "Output Directory" in the left pane
    - Click on "Add Source Directory and fill out the form as follow

Source Directory for Masking

Source Directory:	/Marketing	Browse
Branch Point:	/	
Destination Directory:	/MarketingMaskOut	

Save   Close

- Create new Hadoop task (as in previous steps) by clicking Hadoop -> Task -> New task
- Define the task as shown below by following these steps
- Task Name: **StructuredMasking**
- Task Description: **StructuredMasking**
- Task Type: **Masking** (from the drop down)
- Full Scan: **YES (check box)**
- Structured: **check the box** (this will show "Output Column Form" drop down on the right)
- Output Column Form: **Replace**
- Compliance Policy: **Tutorial**
- Select Directories: **/Marketing**
- **Save and Execute**

DATAGUISE

Uncheck the verify flag before pressing 'Save and Execute'

Review Results

- When the task is finished, the result is displayed in the bottom pane
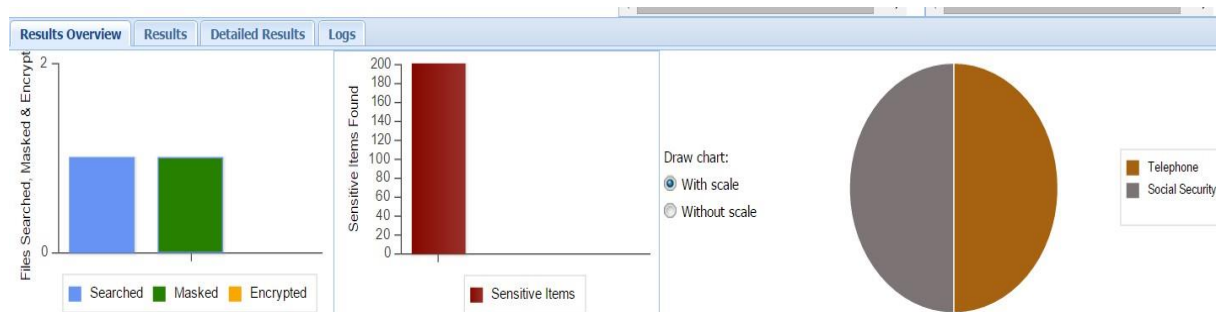
Detailed Results:

- Click on the "Detailed Results" tab on top left of the bottom pane. This will show the details of the scan, which includes Sensitive Data Type, location as well as protection status.

Results Overview	Results	Detailed Results	Logs							
								Clear Filters	Save Results to File	Save Results to Pd
Task Name	Sensitive Data Type ▲	File Path		File Type	Discovery Ty	Content Read	Hit Count	Masked	Encrypted	Increme
StructuredMasking	Social Security # (Digits Only)	//cl29104@st1cl622i29104.blob.core.windows.net/Marketing/payroll.csv					100	true	false	false
StructuredMasking	Telephone (Digits Only)	//cl29104@st1cl622i29104.blob.core.windows.net/Marketing/payroll.csv					100	true	false	false
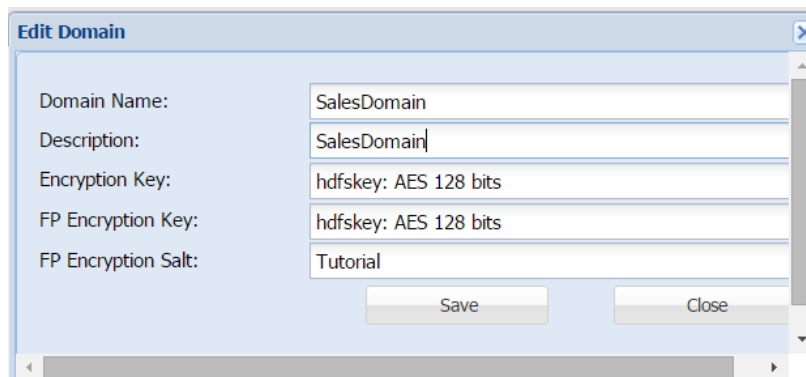
5. Encryption

A domain is needed when a data set requires encryption or decryption. A domain comprises a name, encryption key, one or more destination directories, and one or more policies. Having a domain enables easier management of encryption keys, policies, and ensures unique encryption between different domains.

Unstructured Encryption

Setup a Domain

- Click Domain Definition.
- Click Add Domain.
- Domain Name: **SalesDomain**
- Description: **SalesDomain**
- Encryption Key (choose): hdfskey: **AES 128 bits**
- FP Encryption Key: hdfskey: **AES 128 bits**
- FP Encryption Salt: **Tutorial**
- (FP Encryption Salt is used to maintain consistent encryption results through multiple domains and directories; when the same encryption salt is used, the encrypted values will be the same as the previous ones used with that salt, however if a new salt is used, the values will change)
    - o NOTE: FPE Key and Salt are not used for Standard Encryption
- Click **Save.**

- **Highlight the domain you just created**, (click on that 1 entry at the top to highlight it) and in the middle pane click the checkbox next to the **Tutorial policy** and click on "**Save Policy to Domain**"

- On the bottom pane, click on "Add Directory" to associate a directory to the domain
- Fill out the information as follows and then click save

.

- Now we are ready to create and execute Encryption tasks

Run Unstructured Encryption

- Click Tasks under Hadoop in the left pane
- Click **New Task**.
- Task Name: **UnstructuredEncryption**
- Task Description: **UnstructuredEncryption**
- Task Type: **Field Encryption**
- Compliance Policy: **Tutorial**
- Click "**Select Directories**" button on the right
- Select the directories and/or files that you would like to encrypt from the file selection in the left panel. All files/directories selected will be listed in the bottom right portion of the page.
  Select /Sales in this example. It has 1 unstructured file called charter_one.txt. Click that 1 file.
  Click **Done**
- Click **Save and Execute**.

Results Overview:

Detailed Results:

Clear Filters | Save Results to File | Save Results to P

Task Name	Sensitive Data Type ▲	File Path	File Ty	Discove	Conten	Hit Count	Masked	Encrypted	Incremental S
UnstructuredEncryption	Social Security # (Dash Separation)	//cl29104@st1cl622i29104.blob.core.windows.net/Sales/charter_one.txt				1	false	true	false
UnstructuredEncryption	Social Security # (Digits Only)	//cl29104@st1cl622i29104.blob.core.windows.net/Sales/charter_one.txt				1	false	true	false
UnstructuredEncryption	Social Security # (Space Separation)	//cl29104@st1cl622i29104.blob.core.windows.net/Sales/charter_one.txt				1	false	true	false
UnstructuredEncryption	Telephone (Dash Separation)	//cl29104@st1cl622i29104.blob.core.windows.net/Sales/charter_one.txt				3	false	true	false
UnstructuredEncryption	Telephone (Digits Only)	//cl29104@st1cl622i29104.blob.core.windows.net/Sales/charter_one.txt				3	false	true	false
UnstructuredEncryption	Telephone (Space Separation)	//cl29104@st1cl622i29104.blob.core.windows.net/Sales/charter_one.txt				3	false	true	false

- ▪ A sample BEFORE and FILE CREATE AFTER ENCRYPTION file image below shows an unstructured text file – charter_one.txt. This file contains numerous sensitive items, names, address, telephone numbers, emails, SSN and a custom element. The image below shows the before and after image of the encrypted file – in this case – encrypting all sensitive items – not just SSN and telephone number in our test above.

BEFORE   **FILE CREATED AFTER ENCRYPTION**

Structured Encryption

Setup Structured Data

- Use the same "Payroll" structure setup in the earlier masking step
- Go to Domain Definition → Click Add Domain and fill out the info as follows

Edit Domain ✕

Domain Name:	MarketingDomain
Description:	MarketingDomain
Encryption Key:	hdfskey: AES 128 bits
FP Encryption Key:	hdfskey: AES 128 bits
FP Encryption Salt:	Tutorial
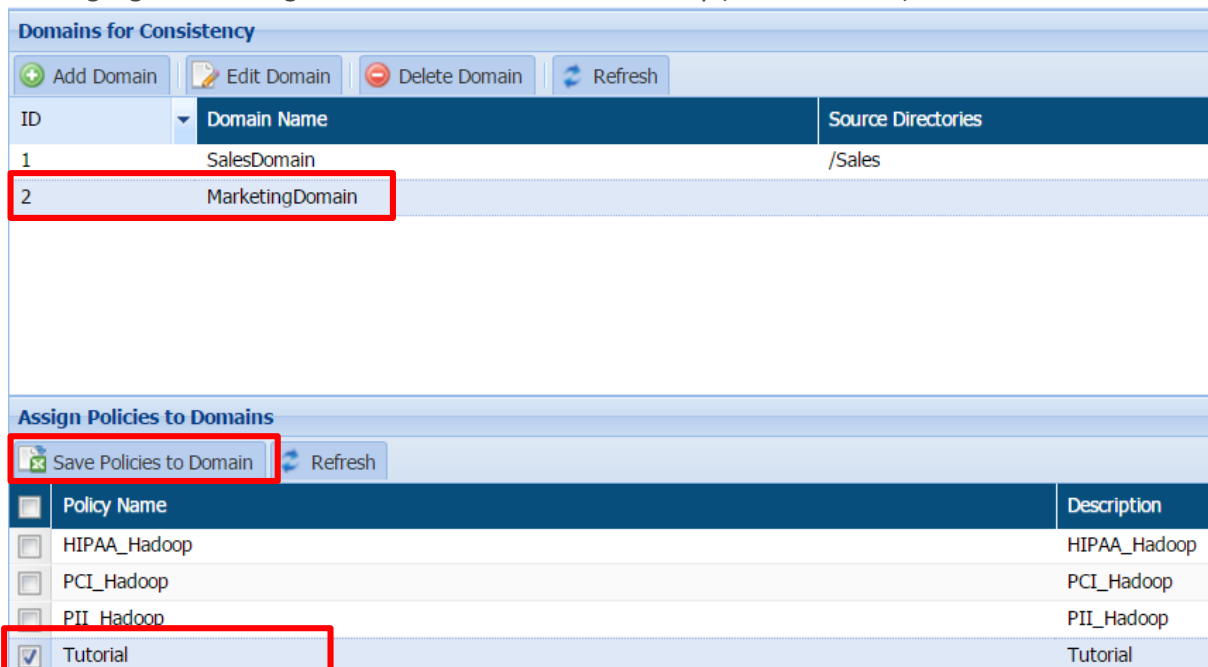
Save   Close

- Also highlight MarketingDomain and click on Tutorial Policy (middle section).

Domains for Consistency

⊕ Add Domain   📝 Edit Domain   ⊖ Delete Domain   🔄 Refresh

ID ▾	Domain Name	Source Directories
1	SalesDomain	/Sales
2	MarketingDomain	

Assign Policies to Domains

🖹 Save Policies to Domain   🔄 Refresh

☐	Policy Name	Description
☐	HIPAA_Hadoop	HIPAA_Hadoop
☐	PCI_Hadoop	PCI_Hadoop
☐	PII_Hadoop	PII_Hadoop
☑	Tutorial	Tutorial

-
- Click '**Save Policies to Domain**'
-
- On the bottom pane, click on "**Add Directory**" to associate a directory to the domain
- Fill out the information as follows and then click **Save**

DATAGUISE

The screen should look similar to this.

- Now we are ready to create and execute another Encryption task.

Run Structured Encryption Job (Field Level)

In this example we will use the structure we created for the payroll file earlier, and encrypt the CCN and SSN.

- Click Tasks tab under the Hadoop on the left pane
- Click **New Task**.
- Task Name: **StructuredEncryption**
- Task Description: **StructuredEncryption**
- Task Type: **Field Encryption**
- Structured: **Click checkbox**
- Compliance Policy: **Tutorial**
- Click "**Select Directories**" on the right and select the **Marketing** directory – click '**Add**'

- Click **Done** once selected to the bottom pane.
- Click "Save and Execute"

Uncheck the verify flag before pressing 'Save and Execute'

Result Structured Encryption job (Field Level) Results

Overview

Detailed Results

Task Name	Sensitive Data Type ▲	File Path	File Type	Discovery Type	Content Rea	Hit Count	Masked	Encrypted	Increm
StructuredEncryption	Social Security # (Digits Only)	//cl29104@st1cl622i29104.blob.core.windows.net/Marketing/payroll.csv				100	false	true	false
StructuredEncryption	Telephone (Digits Only)	//cl29104@st1cl622i29104.blob.core.windows.net/Marketing/payroll.csv				100	false	true	false

DATAGUISE

As a quick check, you should check on the remaining time left in your evaluation:

Trial In-progress

Remaining

| 0 | 1 | 43 | 57 |
| DAYS | HOURS | MINUTES | SECONDS |

Your DgSecure for HDInsight test drive is ready! Follow the instructions in the user guide and use the credentials sent to your email and the environment window to get started.

Stop
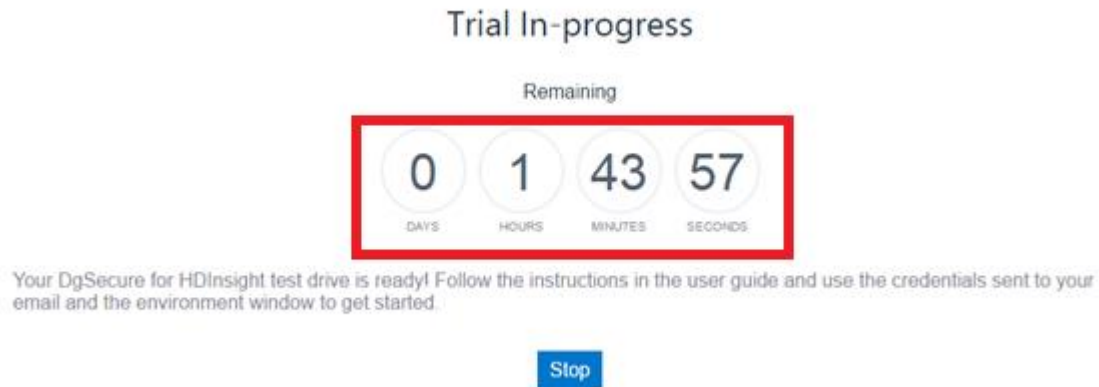
6. Bulk Decryption

Roles and ACL Management

In order to do a decryption task, we first need to add a user and assign sensitive elements that the user has privileges to decrypt.

Setup Role Management

- Click Role Management in the left pane

  ● **Role Management**
  ACL Management
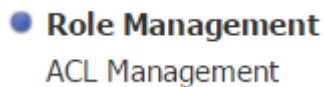
- Click Add Role

Create/Edit Hadoop Role

Role Name: decrypt_role

Role Description: decrypt_role

Cancel   Save

- Role Name: decrypt_role (same for Role Description)
- Press **Save**

- Expand "Edit Sensitive Data Types Access Permissions" middle of screen.

- Press 'Refresh' to populate

Edit Sensitive Data Types Access Permissions

Name	Access
IP Address	☐
Names	☐
NPI	☐
Social Security	☑
Telephone	☐
URL	☐

Click Only Social Security – Not Telephone
- Click **Save**
- Expand "Edit Days Permissions"
- This will set the days and times that decryption can happen for the role
- Click on **Set Time**

Edit Time

Start Time:	End Time:		
00:00:01	23:59:59	☑ Full Day	Add

Remove Column

Start Time	End Time
00:00:01	23:59:59

Cancel · Save

-
- Check '**Full Day'** for each of the days Mon – Fri, and press **Save**

Edit Days Permissions

Set Time · Refresh

Day Name	Time Intervals
Thu	00:00:01-23:59:59
Tue	00:00:01-23:59:59
Fri	00:00:01-23:59:59
Mon	00:00:01-23:59:59
Wed	00:00:01-23:59:59
Sun	

-

Setup ACL's for User

- Click **ACL Management** in the left pane

  ● **ACL Management**

- The Role name 'decrypt_role' should be present at top
- Click on **Refresh** in the 'Groups/Users List'
- Click on User Name: **hdfs**

Groups/Users List

| Save | Refresh |

- ▷ ☐ 📁 root
- ▷ ☐ 📁 sys
- ▷ ☐ 📁 uucp
- ▷ ☐ 📁 lp
- ▷ ☐ 📁 adm
- ◢ ☑ 📂 hdfs
  - ☑ 📄 hdfs
- ▷ ☐ 📁 landscape
- ▷ ☐ 📁 netdev
- ▷ ☐ 📁 plugdev
- ▷ ☐ 📁 nogroup
- ▷ ☐ 📁 dip
- ▷ ☐ 📁 proxy
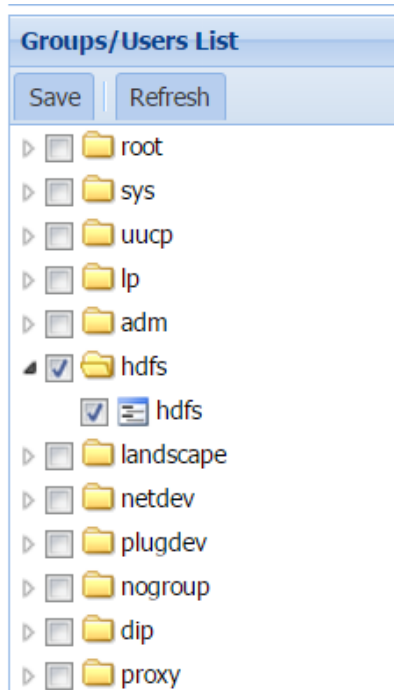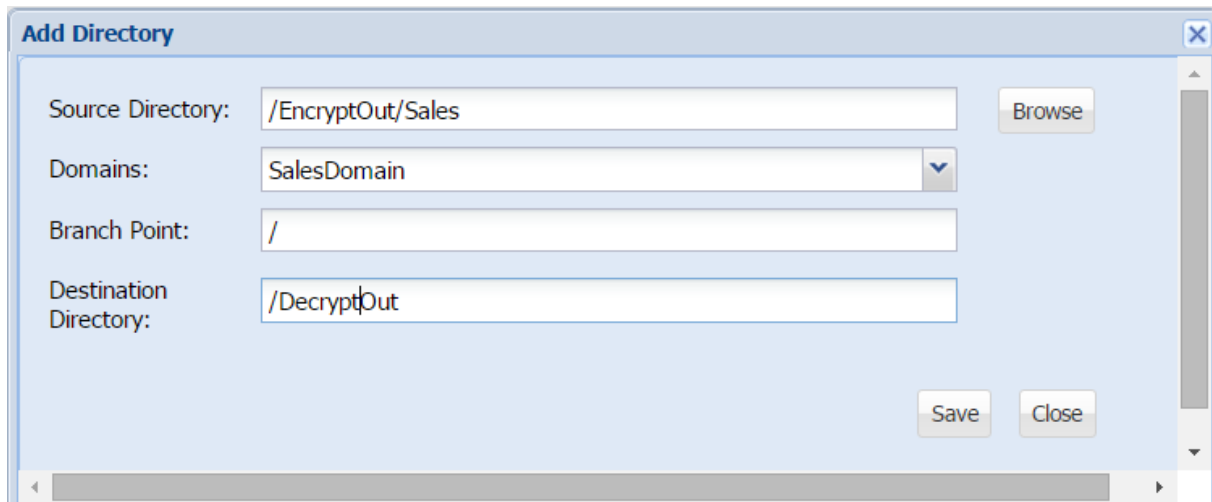
- Click **Save**
- **The user hdfs is running the MapReduce jobs in this evaluation test if Azure.**
- **We have allowed the user (hdfs) the ability to decrypt SSN only.**

Update Encryption Domain to include Decryption

We need to update the domain we defined earlier to associate the directory of the encrypted file, and where we want the decrypted file to go

NOTE: Same domain needs to be defined for Decryption as was used for Encryption for the same set of files

- Click **Domain Definition**.

- Click **Add Directory (bottom portion of screen)**

-
- NOTE: Remember in the Unstructured Encryption task earlier - we chose that the /Sales directory target for encryption was the **/EncryptOut** directory. That means that the /Sales directory was created under the /EncryptOut dir.   We are now using that as the source directory to test Decryption. So the source directory is **/EncryptOut/Sales**
- The output of this task will go to the **/DecryptOut** directory.
- Click **Save**

- The Domain Definition → lower page should look like this:

Source Directory	Domain	Branch Point	Destination Directory
/EncryptOut/Sales	SalesDomain	/	/DecryptOut
/Sales	SalesDomain	/	/EncryptOut
/Marketing	MarketingDomain	/	/StructuredEncryptOut

Run Decryption Job

- Click **HDFS → Tasks** (under Hadoop tab on left)
- Click **New Task**
- Task Name: **Decryption**
- Task Decryption: **Decryption**
- Task Type: **Decryption**
- **Do not check Structured**
- Compliance Policies: **Tutorial**
- Select Directories: **/EncryptOut/Sales**
- Click **Save and Execute**

Results for Decryption Job

The decrypted file can be found under the /DecryptOut directory on the cluster

This evaluation does not allow access to the cluster hadoop behind the scenes to check these files. This test is purely to help train the new user how to navigate within DgSecure and setup tasks for Discovery, Masking, Encryption and Decryption.

Deeper tests and evaluations can be setup within your own Azure cloud. Please contact Dataguise Sales for more information.

7. Reporting

Dashboard

Dashboard provides a rolled up report of all task that have been run and break down of where the sensitive elements were found.

- Click on the "Report" tab from the left pane, which will take you to the dashboard tab.

- The top pane shows the "Report" tab from the left pane, which will take you to the dashboard tab.
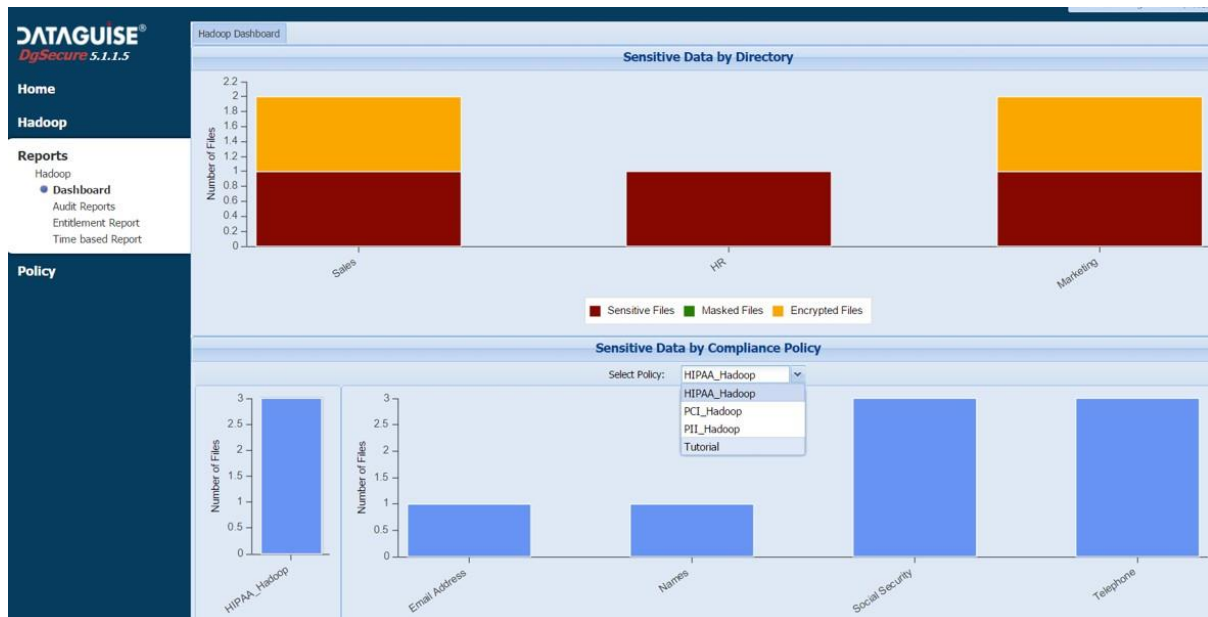
- The bottom pane shows the 3 supplied Policies plus the one you created called 'Tutorial'.
- Select **Tutorial**. This will show the two sensitive items you activated – SSN and Telephone.
- You should now **click on the blue graph above Social Security**
- Click **Update Results** on the lower left of the next screen – if the screens do not populate
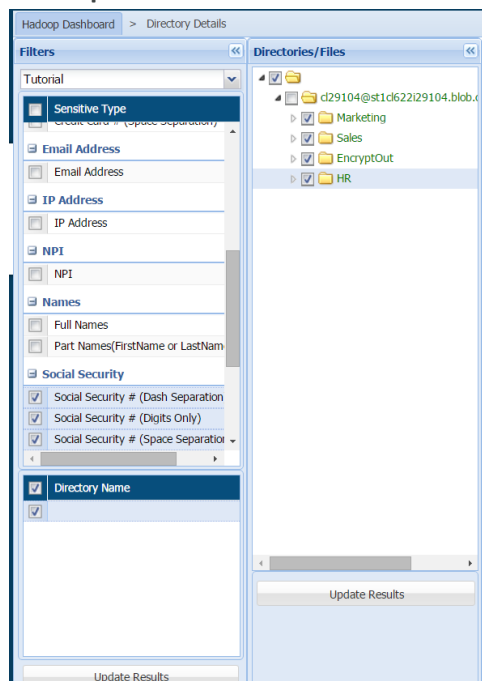
-
- Expand the center pane File browser to show all the subdirectories
- Check all of the subdirectories and click on **Update Results** (bottom of that pane)
- This will repopulate the right-most screen.

- This right-most pane shows every file in the selected directories that contain SSN. It also shows the counts of each, plus whether that file has been masked or encrypted.
- From this single security pane, a Security officer has the ability to check on every single sensitive element within their cluster and detect where any exposure exists. Very powerful.

A PDF or a consumable report can also be downloaded showing this information.

Entitlement Report

Entitlement reporting shows

- ▪ Users configured on the cluster
- ▪ Directories scanned
- ▪ Protection status of the directories
- ▪ User permission for the directories

This helps in triangulating the information about sensitive directories in the cluster and who has what type of access to these directories and files.

- ▪ Click on Entitlement Report
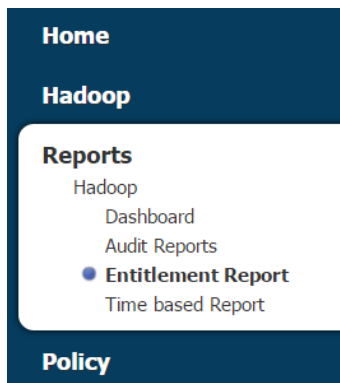
Initial Screen: **press Refresh if the screens are not populated**
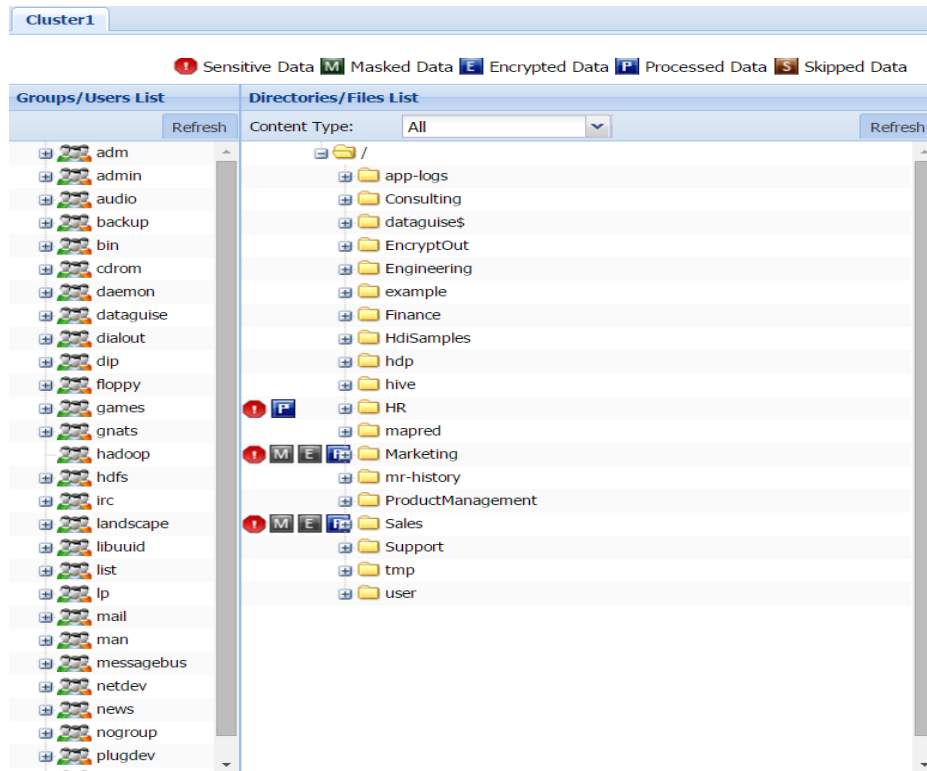
- Expand the user hdfs on left screen
- Also expand the /HR directory in the right screen
- Right-click on user 'hdfs' or 'mapr' and select '**Get Permissions for Selected Directory'**

- Permissions for that specific dir and files within it are displayed

- You can also get the sensitive items found in each file, but expanding the directory and right click on the desired file.
- Right-click on the file /HR/charter_one.txt → Select **'Get Sensitive Type Details For File'**
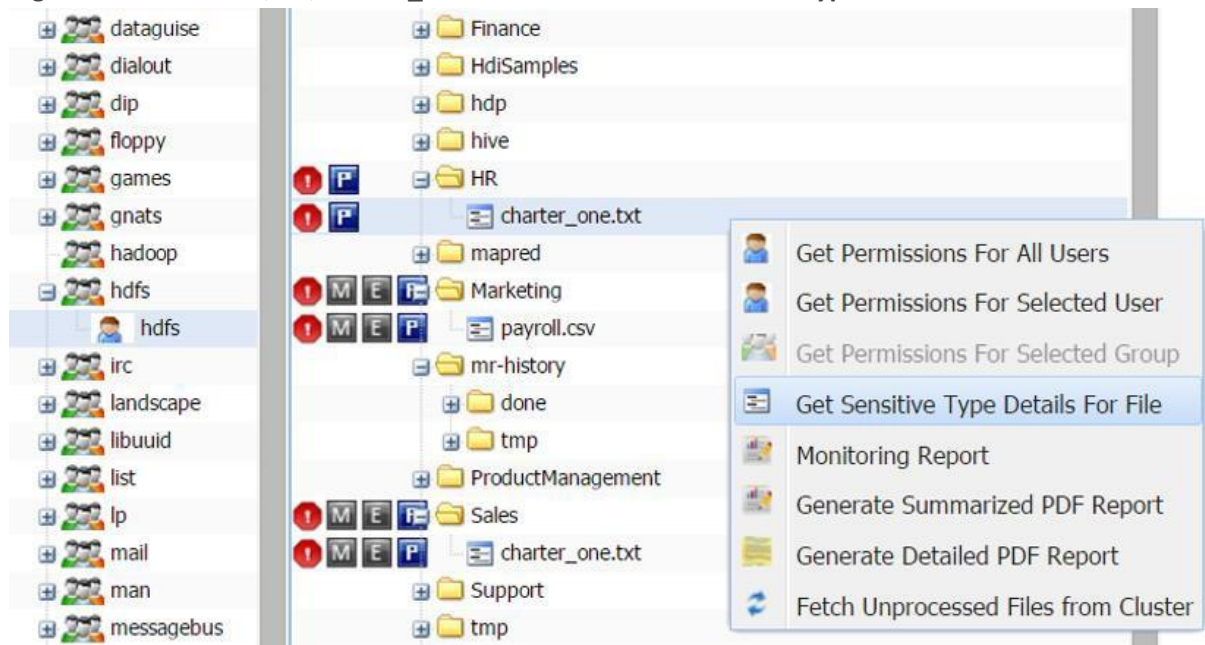
- This shows that Dataguise understands where all of the sensitive data exists within Hadoop. Not only what files have sensitive data but how many of each sensitive element exists within each file. Very powerful.

Sensitive Data

Regex Name	Regex Count	Encrypted	Discovered	Masked
Email Address	6	false	true	false
Part Names(FirstName or LastName)	12	false	true	false
Telephone (Standard)	3	false	true	false
Telephone (Digits Only)	3	false	true	false
Telephone (Dash Separation)	3	false	true	false
Address (Unstructured data only)	2	false	true	false
Telephone (Space Separation)	3	false	true	false
Telephone (Standard without spaces)	3	false	true	false
Social Security # (Space Separation)	1	false	true	false
Social Security # (Digits Only)	1	false	true	false
Social Security # (Dash Separation)	1	false	true	false
Full Names	1	false	true	false

Cancel

Audit

Audit reports keeps track of the tasks submitted through the DgSecure controller and details about the tasks. This information can be viewed in several ways, e.g. time range or event based.

Click on the user Dataguise to see everything that user has done.

Time Based

Time based report shows Searched, Masked and Encrypted files during a particular time period and number of sensitive hits for each policy filtered by date.