

If you analyze your  
Internet of Things data  
in Hadoop...

... you need  
SecureSequester



If you analyze your  
Internet of Things  
data in Hadoop  
you need  
SecureSequester

Your Data Lake is the ideal environment to analyze data you obtain from Internet of Things (IOT) devices you have employed to

But...

Your Data Lake lacks some of the characteristics you have become accustomed to in other environments

- Descriptive information about the contents of your data lake

- Security characteristics you have become accustomed to

If you have personally identifiable information in your data lake to better understand who is using your IOT devices, then you have reputational and regulatory issues that will force you to take action

# BigDataRevealed's SecureSequester employs 3 components

BigDataRevealed employs three components to protect streams of data made available by IoT and other devices that write in your data lake

**The producer is the management workbench used to configure the streams that should be interrogated by BigDataRevealed's SecureSequester**

**The patterns in the streams are identified and false positives are eliminated, leaving what is potentially exposed Personal Identifiable Information (PII)**

**The SecureSequester capabilities of BigDataRevealed's Intelligent Catalog then go to work to encrypt PII data and sequester the keys necessary to reveal PII values**

# The Three Components

1. Producer is used to register potential streams of information to BigDataRevealed. Connections are automatically generated.

Create Producer

Producer Name

Meeting

Topic

stream

Partitions

1

Replication

1

Url

http://stream.meetup.com/2/rsvps

Delimiter

,

Cancel

Create

## 2. The SecureSequester

Administrative workbench is used to define PII patterns and known false positives. Set Duration and parameters for the Stream Job.

Stream Encryption Job

Select Producer

Meeting

Delimiter (default ',')

,

Duration

2

Days

Regular Expression Name

EmailDetailed

Regular Expression

[a-z0-9]+[+\_-][a-z0-9]+(0.2)@[?!.a?]{cdefghilmnoqrstuwxz}[ero]([?!.p(s)a])\b(?:[abdefghijkmnorstvwyz]{c}|c(?:[acdfghiklmnorvxyz]{a}|a(?:[c?!.m(c)p])|d(ejkmoxz)|e(?:[ceghrstu]|du|t|ijkmor))

Add

Regular Expressions

Name	Expression	Action
Local Security Number Valid Numbers Only	([a-z0-9_!-+&@]{1,25} [a-z]{1,25})	

3. SecureSequester will interrogate streams as they are introduced to the data lake, encrypt potential PII, dispatch alerts for review and sequester encrypted source data.

"public"	"yes"	0.0	0/TYEc2zKh/vnVmccgaB6w==	xvMR/qvVFjYNNWukETGs0nmQJbTqnmLUs6uH4ZpnfS0=	"Priority Fitness"	-94.702271
"public"	"yes"	0.0	0/TYEc2zKh/vnVmccgaB6w==	xvMR/qvVFjYNNWukETGs0nmQJbTqnmLUs6uH4ZpnfS0=	"Priority Fitness"	-94.702271
"public"	"yes"	0.0	BIOPaSRnLUelsVyO47Dw==	+ZiZOndalwRGLz9cEiFS3fw==	"Innisfree Garden"	-73.750755
"public"	"yes"	0.0	BIOPaSRnLUelsVyO47Dw==	+ZiZOndalwRGLz9cEiFS3fw==	"Innisfree Garden"	-73.750755
"public"	"yes"	0.0	BIOPaSRnLUelsVyO47Dw==	+ZiZOndalwRGLz9cEiFS3fw==	"Innisfree Garden"	-73.750755
"public"	"no"	0.0	R+sPacB87LUQIKRUYbVA==	rLkZzHBQwmLp/WXf7SswJbTqnmLUs6uH4ZpnfS0=	1.24545332E8	06f4iW3NAJcSoGkYGI3TVXJHfKvqvGuZGIYn
"public"	"no"	0.0	R+sPacB87LUQIKRUYbVA==	rLkZzHBQwmLp/WXf7SswJbTqnmLUs6uH4ZpnfS0=	1.24545332E8	06f4iW3NAJcSoGkYGI3TVXJHfKvqvGuZGIYn
"public"	"no"	0.0	R+sPacB87LUQIKRUYbVA==	rLkZzHBQwmLp/WXf7SswJbTqnmLUs6uH4ZpnfS0=	1.24545332E8	06f4iW3NAJcSoGkYGI3TVXJHfKvqvGuZGIYn
"public"	"no"	0.0	R+sPacB87LUQIKRUYbVA==	rLkZzHBQwmLp/WXf7SswJbTqnmLUs6uH4ZpnfS0=	1.24545332E8	06f4iW3NAJcSoGkYGI3TVXJHfKvqvGuZGIYn
"public"	"yes"	0.0	zwLac4tbelMxUGTbVwBKQ=	OZstCKt5GYfFz9GIWKP82+w==	"Round Valley Recreation Area Boat Launch"	-74.905889
"public"	"no"	0.0	R+sPacB87LUQIKRUYbVA==	rLkZzHBQwmLp/WXf7SswJbTqnmLUs6uH4ZpnfS0=	1.24545332E8	06f4iW3NAJcSoGkYGI3TVXJHfKvqvGuZGIYn
"public"	"no"	0.0	R+sPacB87LUQIKRUYbVA==	rLkZzHBQwmLp/WXf7SswJbTqnmLUs6uH4ZpnfS0=	1.24545332E8	06f4iW3NAJcSoGkYGI3TVXJHfKvqvGuZGIYn
"public"	"no"	0.0	R+sPacB87LUQIKRUYbVA==	rLkZzHBQwmLp/WXf7SswJbTqnmLUs6uH4ZpnfS0=	1.24545332E8	06f4iW3NAJcSoGkYGI3TVXJHfKvqvGuZGIYn
"public"	"no"	0.0	R+sPacB87LUQIKRUYbVA==	rLkZzHBQwmLp/WXf7SswJbTqnmLUs6uH4ZpnfS0=	1.24545332E8	06f4iW3NAJcSoGkYGI3TVXJHfKvqvGuZGIYn

## I-o-T as easy as 1-2-3

### Protection for EU GDPR – US and worldwide Data Protection, Sequester / Encryption

BigDataRevealed employs three components to protect streams of data made available by IOT and other devices that write in your Data Lake

1. Producer is used to register potential streams of information to BigDataRevealed. Connections are automatically generated.

#### 2. The SecureSequester

Administrative workbench is used to define PII patterns and known false positives. Set Duration and parameters for the Stream Job.

3. SecureSequester will interrogate streams as they are introduced to the data lake, encrypt potential PII, dispatch alerts for review and sequester encrypted source data.

#### Create Producer

Producer Name

Meeting

Topic

stream

Partitions

1

Replication

1

Url

http://stream.meetup.com/2/rsvps

Delimiter

.

Cancel

Create

#### Stream Encryption Job

Select Producer

Meeting

Delimiter (default '.')

.

Duration

2

Days

Regular Expression Name

EmailDetailed

Regular Expression

[a-z0-9]+([+\_]([a-z0-9]+)(0,2)@)?\.[a-z]?([cdefghimnoqrstuvxyz]?ero)([?rp]s[a])b([?abdefghijkmnorstvwyz]?c)?([acdfghiklmnorstvwyz]?at)([?mop])([de]kmoz)?e([?ceghrstu])du([?ij]kmor)?g

Add

Regular Expressions

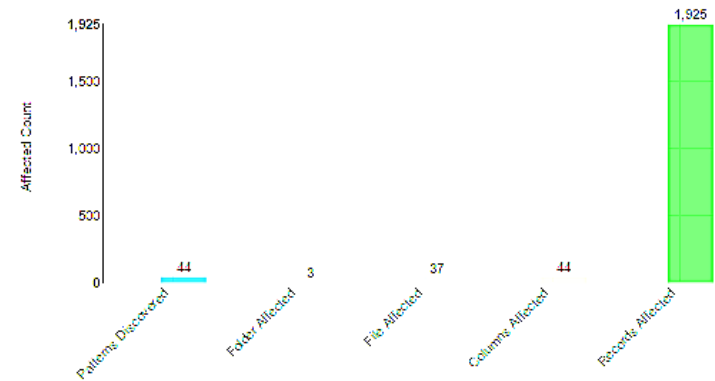
Name	Expression	Action
Social Security Number Valid Numbers Only	((a-z0-9_\-+)([de-z1-+])([a-z1](2,5))	

"public"	"yes"	0.0	0/TYEcZzKhVnVmccgaB6w==	xvMRNqVFjYNNWuE7Gs0nmQUbTqnnUs6tuH2pnfs0=	"Priority Fitness"	-94.702271
"public"	"yes"	0.0	0/TYEcZzKhVnVmccgaB6w==	xvMRNqVFjYNNWuE7Gs0nmQUbTqnnUs6tuH2pnfs0=	"Priority Fitness"	-94.702271
"public"	"yes"	0.0	BIOPaSrRnUJelsVyOt47Dw==	+2IZOndalwRGLz9cEIS3fw==	"Innisfree Garden"	-73.750755
"public"	"yes"	0.0	BIOPaSrRnUJelsVyOt47Dw==	+2IZOndalwRGLz9cEIS3fw==	"Innisfree Garden"	-73.750755
"public"	"yes"	0.0	BIOPaSrRnUJelsVyOt47Dw==	+2IZOndalwRGLz9cEIS3fw==	"Innisfree Garden"	-73.750755
"public"	"no"	0.0	R+sPacB87LuUQIKRUYbVA==	rLkZ2HhBQwmlP/WX7FSswJbTqnnUs6tuH2pnfs0=	1.24545332E8	064fW3NAJcSoGkYGI3TVXJfKvqyGu2GIYn
"public"	"no"	0.0	R+sPacB87LuUQIKRUYbVA==	rLkZ2HhBQwmlP/WX7FSswJbTqnnUs6tuH2pnfs0=	1.24545332E8	064fW3NAJcSoGkYGI3TVXJfKvqyGu2GIYn
"public"	"no"	0.0	R+sPacB87LuUQIKRUYbVA==	rLkZ2HhBQwmlP/WX7FSswJbTqnnUs6tuH2pnfs0=	1.24545332E8	064fW3NAJcSoGkYGI3TVXJfKvqyGu2GIYn
"public"	"no"	0.0	R+sPacB87LuUQIKRUYbVA==	rLkZ2HhBQwmlP/WX7FSswJbTqnnUs6tuH2pnfs0=	1.24545332E8	064fW3NAJcSoGkYGI3TVXJfKvqyGu2GIYn
"public"	"yes"	0.0	zwLarS4tbelMxUGTbvBtKQ==	OZstCk15GYfz9GWP82+w==	"Round Valley Recreation Area Boat Launch"	-74.909889
"public"	"no"	0.0	R+sPacB87LuUQIKRUYbVA==	rLkZ2HhBQwmlP/WX7FSswJbTqnnUs6tuH2pnfs0=	1.24545332E8	064fW3NAJcSoGkYGI3TVXJfKvqyGu2GIYn
"public"	"no"	0.0	R+sPacB87LuUQIKRUYbVA==	rLkZ2HhBQwmlP/WX7FSswJbTqnnUs6tuH2pnfs0=	1.24545332E8	064fW3NAJcSoGkYGI3TVXJfKvqyGu2GIYn
"public"	"no"	0.0	R+sPacB87LuUQIKRUYbVA==	rLkZ2HhBQwmlP/WX7FSswJbTqnnUs6tuH2pnfs0=	1.24545332E8	064fW3NAJcSoGkYGI3TVXJfKvqyGu2GIYn
"public"	"no"	0.0	R+sPacB87LuUQIKRUYbVA==	rLkZ2HhBQwmlP/WX7FSswJbTqnnUs6tuH2pnfs0=	1.24545332E8	064fW3NAJcSoGkYGI3TVXJfKvqyGu2GIYn

Cataloguing / Metadata / Columnar Naming

1. Executive Summary of Discovery Process and Patterns Detected.

Pattern Discovery



Column Classification

Business Classification	Total Files	Total Columns	Total Records
NA	53	82	598
EMAIL_PATTERN	26	9	74
SSN	16	4	49
US_CREDIT_CARDS	11	1	39
CREDITCARD	11	1	39
TAX_PAYER_ID	16	2	43
IP_ADDRESS	18	8	53
DATE_TIME	7	16	21

2. BigDataRevealed creating Business Column Classification.

Column Name/ID	Business Classification	Percent	Action
HDFS Col Pos	NA		
Last Name	NA		
Email-Address	EMAIL_PATTERN	81.74%	
IP Address	IP_ADDRESS	87.83%	
IP Address	NA		
Credit Card 2	CREDITCARD	34.66%	
Credit Card 2	US_CREDIT_CARDS	35.23%	
Race	NA		
Social Security Number	TAX_PAYER_ID	2.61%	
Social Security Number	SSN	70.43%	

1 - 10 of 10

1025

3. Creating the File Headers, Catalog Info and Collaborative usage.

File Content

File Path: /sampledata/MOCK\_DATA.csv

Select Option

Add/Update File Column

Select Column Headers from User Input

Import File Header

Export File Header

First Five Row From Latest Modify File Content:

HDFS Col Pos	First Name	Last Name	Email-Address	Race	IP Address	Mobile	Credit Card 2	Gender	Race	Social Security Number
id	first_name	last_name	email	country	ip_address	phone_number	credit_card	gender	race	MEICKFuXRTq7wPu3YTFBQ==
1	Lisa	Jenkins	ljenkins0@arstechnica.com		183.32.221.138	3-(077)695-6786	4508506137315268		Indonesian	YEvNbEGJUGORNIikyg==
2	James	Marshall	jmarshall1@bloomberg.com				67098475265817698	M	Alaska Native	Q4AE+UKFNs5IM+XvKkrczA==
3	Judith	Fowler	jfowler2@tiny.cc	Philippines		0-(751)958-0053	3546183479556523	F	Cheyenne	M1y1kgEXUMtzaJbaYojAg==
4	Michelle	Shaw			166.87.11.96	5-(206)112-9176	30596520921504	F	Cheyenne	rL528xvMtgWQG4NvhuAJw==

# Pattern Discovery of Private Data then Sequester / Encrypt Protection for EU GDPR









1. View File/Column where Personal Data or Discovered Data was found and select Sequester

Pattern Discovery Results

File Path: /sampledata/MOCK\_DATA.csv  
pattern: Social Security Number

Q Search

ConsolidatorSequester

Column Name/ID	File Name	Discovery Pattern	Action
First Name	/sampledata/MOCK_DATA.csv	Social Security Number	 
Race	/sampledata/MOCK_DATA.csv	Social Security Number	 
Credit Card 2	/sampledata/MOCK_DATA.csv	Social Security Number	 
Social Security Number	/sampledata/MOCK_DATA.csv	Social Security Number	 

1 - 4 of 4

10

25

50

100









2. Select individual columns by data type to Encrypt, or select entire file to Encrypt. Then Run Process.

Pattern Discovery Results

File Path: /sampledata/MOCK\_DATA.csv  
pattern: Social Security Number

Q Search

ConsolidatorSequester

Column Name/ID	File Name	Discovery Pattern	Action
First Name	/sampledata/MOCK_DATA.csv	Social Security Number	 
Race	/sampledata/MOCK_DATA.csv	Social Security Number	 
Credit Card 2	/sampledata/MOCK_DATA.csv	Social Security Number	 
Social Security Number	/sampledata/MOCK_DATA.csv	Social Security Number	 

1 - 4 of 4

10

25

50

100

3. View the results of what Columns or if the entire file is encrypted by seeing the actual data.

First Five Row From Latest Modify File Content:

HDFS Col Pos	First Name	Last Name	Email-Address	Race	IP Address	Mobile	Credit Card 2	Gender	Race	Social Security Number
1	Lisa	Jenkins	ljenkins0@arstechnica.com		183.32.221.138	3-(077)695-6786	4508506137315268		Indonesian	YEvN8EG3Ji3G0RINJlkyg==
2	James	Marshall	jmarshall1@bloglines.com				67098475265817698	M	Alaska Native	Q44E+UKFNs5Wt+X/KKrczA==
3	Judith	Fowler	jfowler2@tiny.cc	Philippines		0-(751)958-0053	3546183479556523	F		M1yIlgEkUMbzaJbwYgAg==
4	Michelle	Shaw			166.87.11.96	5-(206)112-9176	30596520921504	F	Cheyenne	rL528XvMhgWQGz4tVHuAJw==



# Pattern Discovery of Private Data then Consolidate into one Folder for further Analysis and Remediation









1. View File/Column where Personal Data or Discovered Sensitive Data was found and select Sequester / Encrypt the sensitive data.

Pattern Discovery Results

File Path: /sampledata/MOCK\_DATA.csv  
pattern: Social Security Number

Search

Consolidator Sequester

Column Name/ID	File Name	Discovery Pattern	Action
First Name	/sampledata/MOCK_DATA.csv	Social Security Number	 
Race	/sampledata/MOCK_DATA.csv	Social Security Number	 
Credit Card 2	/sampledata/MOCK_DATA.csv	Social Security Number	 
Social Security Number	/sampledata/MOCK_DATA.csv	Social Security Number	 

1 - 4 of 4









10 25 50 100

2. Select files to be copied into a new folder containing like data for further analytics, and remediation.

Search

/FolderForSocialSecurityFound

Run Cancel

Column Name/ID	File Name	Discovery Pattern	Action
<input type="checkbox"/> First Name	/sampledata/MOCK_DATA.csv	Social Security Number	 
<input type="checkbox"/> Race	/sampledata/MOCK_DATA.csv	Social Security Number	 
<input type="checkbox"/> Credit Card 2	/sampledata/MOCK_DATA.csv	Social Security Number	 
<input checked="" type="checkbox"/> Social Security Number	/sampledata/MOCK_DATA.csv	Social Security Number	 

1 - 4 of 4

10 25 50 100

3. View the results of files that were written into the New Folder as per number 2.

FileWithSocialSecurity

FolderForSocialSecurityFound

MOCK\_DATA.csv

hbase









metadataanalytics

sampledata

STEVESSIN

test

tmp

Column Name/ID	File Name	Discovery Pattern	Action
First Name	/sampledata/MOCK_DATA.csv	Social Security Number	 
Race	/sampledata/MOCK_DATA.csv	Social Security Number	 
Credit Card 2	/sampledata/MOCK_DATA.csv	Social Security Number	 
Social Security Number	/sampledata/MOCK_DATA.csv	Social Security Number	 

1 - 4 of 4

10 25 50 100



GDPR, or the General Data Protection Regulations, mandate you to have your data house in order if you plan on doing business in Europe. Similar regulations can be expected in the USA.

Your data lake, which makes analysis of data from IOT and other sources, leaves gaping holes which make compliance difficult. Luckily, we at BigDataRevealed have an easy to implement solution to fill the gaps.

For more information, email [info@bigdatarevealed.com](mailto:info@bigdatarevealed.com) 847-791-7838

