

Regulators have begun passing and enforcing legislation which protects their citizenry from cyberthreat and mandates a right to be forgotten. GDPR is one of the first and most stringent of these regulations. The EU is set to make examples of companies either unwilling or unable to comply with GDPR, this deck gives a sense of what you must consider doing to comply.

What you must ensure



Productionalized
Discovery



Repeatable,
auditable solution



Enable the
Right to be
Forgotten



Insulate in flight
transactions from
cyber-attacks

ENABLERS

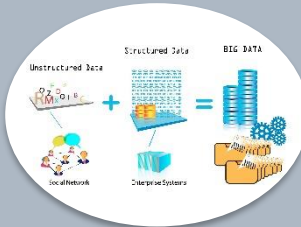
The
Intelligent
Catalog

Single Platform

Functionally rich
rules
maintainable by
non-technicians

Single PII
discovery zone

Road Map to
Compliance
Privacy
Assurance
Calculator



Click on each
process, execute
each process and
have accomplished
your role in GDPR
and Protecting
You're Customer
trusted Private
Information
leaving the hackers
with meaningless
highly Encrypted
Information

BigDataRevealed
Architecture Slide

What must you do to comply with GDPR

Home

1

- Identify information that uniquely identifies a consumer, either directly or indirectly, residing anywhere in your portfolio of files and systems maintained within your four walls, within cloud environments you use, within your backups or within partner environments.

2

- Develop a plan to secure personally identifiable information. Document that plan sufficiently well so that regulators can be satisfied. And finally execute the plan while documenting milestones achieved, again to satisfy regulators

3

- Take the necessary steps to protect personally identifiable information from being stolen or altered inappropriately through a well devised program

4

- Have a facility that is able to locate information for a particular consumer everywhere in your environment so that a request from any consumer to be forgotten can be fulfilled within 48 hours.

5

BigDataRevealed
Architecture Slide

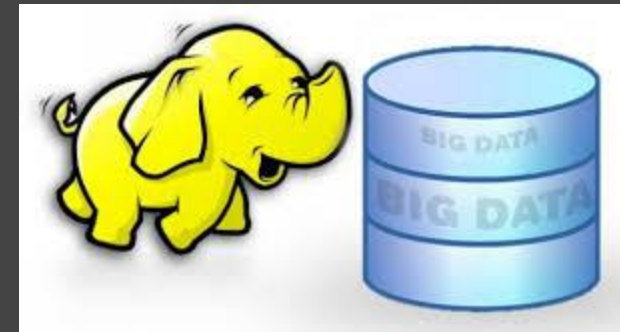
Centralize Legacy, Structured, Unstructured and Live Streaming Data into a file system that supports all data types including binary files such PDF, Office Documents, XML, Pictures and more..

Apache™ Hadoop® Staging ODS

[Home](#)[Next Page](#)

Think about the value of mainframe data, AS/400 data, Oracle, SQL Server, DB2, Teradata, PDF's, Office Documents, Emails and other data being processed by one application using a single platform. Imagine the exponential improvement in BI reporting, demonstration of regulatory preparedness, collaboration among data scientists and the ease of adjusting to new data entering your environment every day.

A new way or addressing data management will be required to keep up with the pace of data growth and complexity, especially in the coming age of data regulation where Personally Identifiable Information PII must be protected, even if pieces of PII are scattered across multiple files of various originations.



You may choose Apache™ Hadoop® over a vendors Hadoop as it is 100% open source and Free. BigDataRevealed offers a completely configured VM with Apache Hadoop, an installation file to run on your current or new Hadoop environment or even on Amazon or other Clouds.

Centralize Legacy, Structured, Unstructured and Live Streaming Data into a file system that supports all data types including binary files such PDF, Office Documents, XML, Pictures and more..

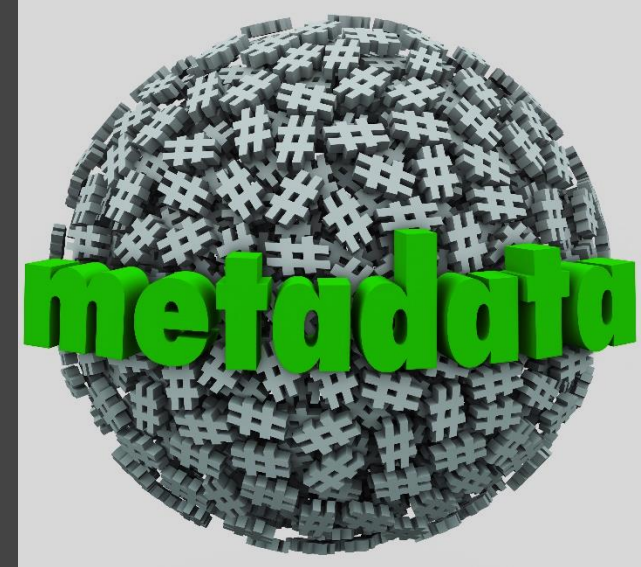
Apache™ Hadoop® Staging ODS

Intelligent Catalog / Metadata

BigDataRevealed's primary advantage over other solutions is the creation of our Intelligent Catalog which provides knowledge of every field in every file and gives the ability to logically join files of varied structures when conducting Indirect Identifier discovery.

We have combined a callable, accurate set of processes that are repeatable and collaborative with a methodology that stores the necessary information in the BDR-Intelligent Catalog Metadata repository. Our Intelligent Catalog and related processes store necessary information developed in earlier steps for use in subsequent processes, thereby avoiding the need to repeat heavy algorithm processing.

You need metadata developed from actual pattern recognition algorithms designed to identify PII that includes Direct and Indirect Identifiers, and you need it developed for all your data no matter what it's source or structure may be.



You may choose Apache™ Hadoop® over a vendors Hadoop as it is 100% open source and Free. BigDataRevealed offers a completely configured VM with Apache Hadoop, an installation file to run on your current or new Hadoop environment or even on Amazon or other Clouds.

Home

Next
Page



Intelligent Catalog / Metadata – Cont.

[Home](#)

We store the metadata necessary for you to respond to regulators and consumers

- The privacy regulations becoming law around the world (GDPR, PrivacyShield, IDP, Cloud-A) call for an ability to prove to regulators that you can administer the right to be forgotten. That means you can identify where a customer's data exists at a moment's notice.
- We create a functionally rich metadata layer as part of the intelligent catalog which we use for what would otherwise be an administrative nightmare.
- We also provide you with the ability to enrich our metadata by augmenting it with metadata from other sources and tools, such as your technical data integration (ETL) tools, your compliance engines, your regulatory reporting engines and any other source you feel aids in your ability to respond to consumers and regulators in meeting the obligation to execute the right to erasure now endowed upon consumers.

Functionally Rich Rules



Big DATA REVEALED
RADAR FOR YOUR DATA LAKE

2270

Modify Global Data

Regex Groups

Search

Group Name

Group 1

ZTTest

Steven Meister

Group 4

PII SSN, Email

Mark

human group

Regex List of ZTTest

Regex Name	Description	Action
Email	Email	
SSN	Social Security #	
SocialSecurityNumber	Social Security Number	
Social Security Number Valid Numbers Only	NULL	
Medicare ID	Medicare ID	
Credit Card Simple Format	Credit Card Simple Format Only	
Credit Card Detailed	Credit Card Majors USA Detailed	
Credit Card CVV	Credit Card CVV	
Individual Taxpayer ID Number	Individual Taxpayer ID Number	

+ New Group

Action

Big DATA REVEALED
RADAR FOR YOUR DATA LAKE

2270

Indirect Identifiers Detail

Search

Group Name

Location Gender Date (Birth Date) IP Group 1

GDPRGroup

Gender

ZipIPBDate

ipemailssn

4values

1 - 6 of 6

Update Indirect Group

Group Name

Location Gender Date (Birth Date) IP Group 1

Description

find files with gender, IP Address and zip code

☐ Telephone USA

☐ Telephone International

☐ Login Name

☐ Screen Name

☐ First Name

☐ Last Name

☐ Country 2 Alpha

☐ Country Code 3 Upper Case

☐ Country Code 3 Num ISO

☐ State Code 2 alpha

☒ Postal Code USA

☒ Gender

☐ Class Grades

☐ Dollar Amount 2 decimals

☒ Dates Detailed

☐ Image Files

☐ HTML Tags

☐ Telephone USA CAN

Cancel

Update

By

Regex Count

Action

4

2

1

3

3

4

10

25

50

100

2017 © BigDataRevealed - v1-0-alpha built on 2017-09-26 1

Quick Discovery

Step One Run Discovery Stats

- Have a means of Identifying Personally Identifiable Information (PII) on File at a moment's notice
- Quickly Searches most vulnerable Personally Identifiable Information
- Displays all found results for each column and the percentage that pattern represents of the total entries.

Share the Intelligent Catalog and Metadata

- View, share and export the results for use in other Catalog / Metadata Repositories
- Users can add or modify Metadata based on users authority

User Collaboration and Sharing and Validating / Updating Metadata

- Adjust columnar business classifications naming to reflect the actual contents of the column.
- Data Scientists and Management (Stewards) will benefit from this accurate columnar naming created by actual data patterns

Home

Next
Page

Quick Discovery

Data Discovery Graphical Viewer with results and drill down into actual domain data values for additional discovery and validation.

Short Video of Discovery



Home

[illegible]

Complete Discovery & Remediation

Short Video of Discovery

[Home](#)[Next Page](#)

If you have any Data Subjects (Consumers, Employees, Suppliers, Partners, Citizens, Patients etc.) sharing their identity with you from Europe or while traveling in Europe, you have to take the new GDPR regulations seriously, which requires you to be able to identify all the information you have about a Data Subject, whether identified by key data items (national insurance number, tax payer id, name, address, credit card numbers, etc.) or indirect information that allows a Data Subject to be identified when multiple fields are grouped together (address, professional affiliations, HIPAA and other medical records or any other information that allows the identification of an individual).

Step One Run Discovery Stats

- Quickly Searches Requested Personally Identifiable Information for Direct Identifiers and Cross File Indirect Identifiers.
- Displays results by each column for every found pattern and what percentage that pattern represents

Share the Intelligent Catalog and Metadata

- View, share and even export the results for use in other Catalog / Metadata Repositories
- Users can add or modify Metadata based on users authority

User Collaboration and Sharing and Validating / Updating Metadata

- Assign the suggested / validated columnar business classification naming to the columnar headings of the files
- Data Scientists and Management (Stewards) will benefit from this accurate columnar naming created by actual data patterns

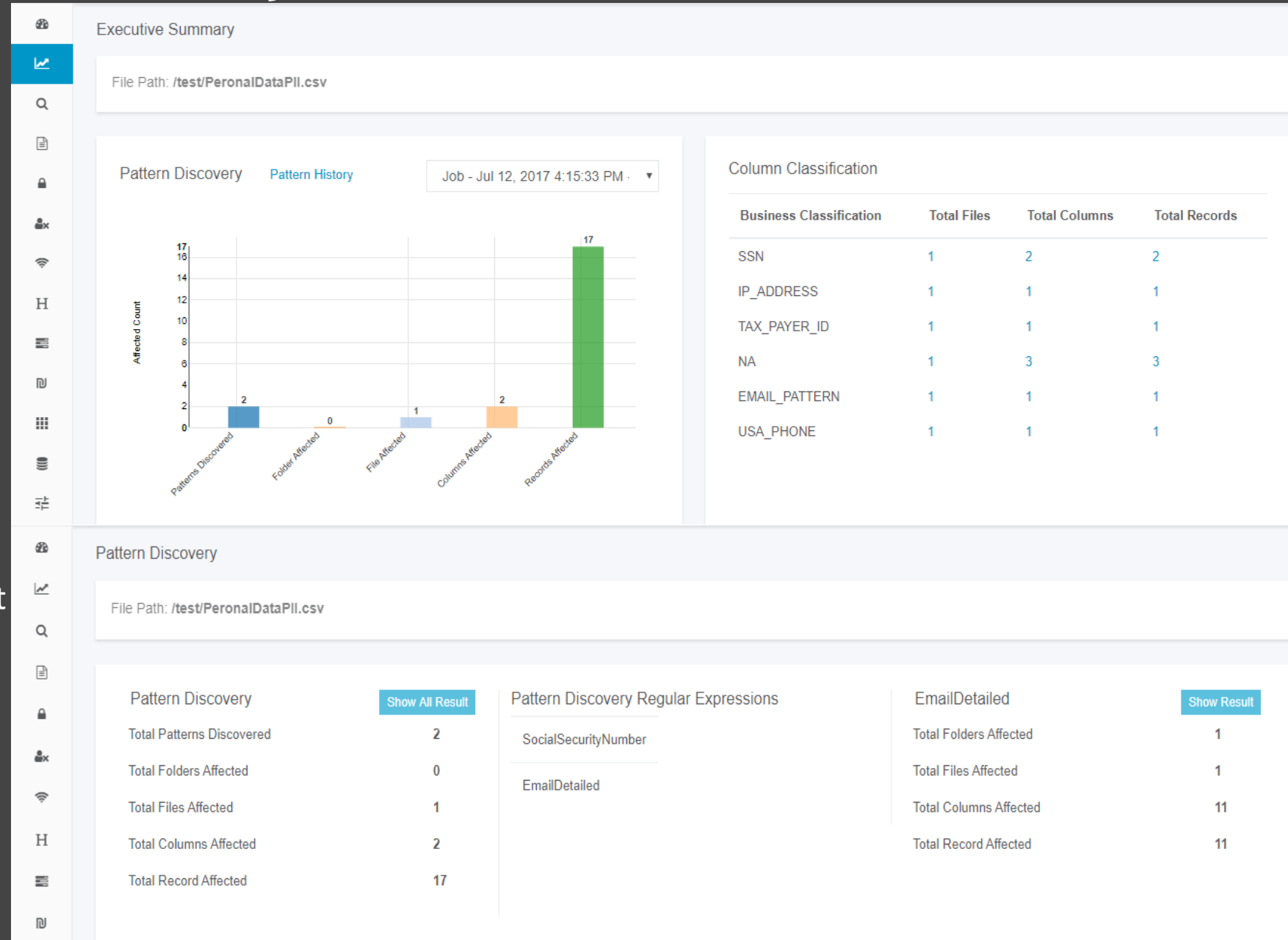
Prepare files and or file columns for Encryption Remediation or full Sequestering of the File to an Encrypted Zone

- Be able to meet the demands of a citizen exercising their GDPR right to be forgotten from all points of identification, whether on your premises, on a backup environment or in a partner's environment storing PII data on your behalf
- Provide a means to protect PII data (that directly or indirectly identifies an individual) from cyber-threats
- Make the choice to either sequester a complete File or to Encrypt one or more Columns of Data in a file.. In the case of Indirect Identifiers and the Right to be Forgotten, you may need to encrypt one or more Column / Row combinations that are found in multiple files.

Complete Discovery

[Home](#)[Next Page](#)
























If you have any Data Subjects (Consumers, Employees, Suppliers, Partners, Citizens, Patients etc.) sharing their identity with you from Europe or while traveling in Europe, you have to take the new GDPR regulations seriously, which requires you to be able to identify all the information you have about a Data Subject, whether identified by key data items (national insurance number, tax payer id, name, address, credit card numbers, etc.) or indirect information that allows a Data Subject to be identified when multiple fields are grouped together (address, professional affiliations, HIPAA and other medical records or any other information that allows the identification of an individual).



Complete Discovery

[Home](#)[Next Page](#)

If you have any Data Subjects (Consumers, Employees, Suppliers, Partners, Citizens, Patients etc.) sharing their identity with you from Europe or while traveling in Europe, you have to take the new GDPR regulations seriously, which requires you to be able to identify all the information you have about a Data Subject, whether identified by key data items (national insurance number, tax payer id, name, address, credit card numbers, etc.) or indirect information that allows a Data Subject to be identified when multiple fields are grouped together (address, professional affiliations, HIPAA and other medical records or any other information that allows the identification of an individual).



Business Classification

File Path: /test/PeronalDataPII.csv

Business Classification

IP_ADDRESS

TAX_PAYER_ID

USA_PHONE

NA

EMAIL_PATTERN

SSN



Pattern SSN

Total Files Affected 1

Total Columns Affected 2

Total Records Affected 2

















Records

Column Name/ID ▾	businessClassification ▾	percent ▾	Action
IP Address	SSN	2.7972	
Social Security Number	SSN	71.8282	

1 - 2 of 2

Consolidator

Sequester

Column Name/ID ▴	File Name ▾	Discovery Pattern ▾	Action
Column Position in HDFS	/test/PeronalDataPII.csv	EmailDetailed	 
First Name	/test/PeronalDataPII.csv	SocialSecurityNumber	 
First Name	/test/PeronalDataPII.csv	EmailDetailed	 
Last Name	/test/PeronalDataPII.csv	EmailDetailed	 
EMAIL_PATTERN	/test/PeronalDataPII.csv	EmailDetailed	 
Country Code	/test/PeronalDataPII.csv	SocialSecurityNumber	 
Country Code	/test/PeronalDataPII.csv	EmailDetailed	 
IP Address	/test/PeronalDataPII.csv	SocialSecurityNumber	 

Home

Data Discovery Results

Server:

Server Port:

File Path: /test/PeronalDataPII.csv

Delimiter: ,

Timestamp: Jul 12, 2017 4:15:01 PM

Metadata	Social Security Reviewed by Data Steward
Description	These are true real live Social Security Numbers and need to be Encrypted
Created By	hadoop - Sep 24, 2017 5:11:43 PM
Updated By	hadoop - Sep 24, 2017 5:11:43 PM

Save
Comment

Column Name/ID	Min Value	Max Value	Mean	Common	Min Length	Max Length	Unique Percent	Domain	Unique Count	Duplicate Count
Social Security Number	NA	NA	NA	<BLANK>	0	11	0.09%	835	834	1

Column Value	Pattern	Classification	Custom Classifications	Frequency	Action
<div style="display: flex; align-items: center; gap: 10px;"> <input type="text" value="Search"/> <div> <input checked="" type="checkbox"/> Case Sensitive </div> <div> <input type="text" value="Search"/> </div> <div> <input type="text" value="Search"/> </div> <div> <input type="text" value="Search"/> </div> <div> Apply </div> </div>					
<BLANK>	<BLANK>		NA	167	✕
108-15-0754	999-99-9999		NA	1	✕
191-04-0961	999-99-9999		NA	1	✕
117-41-5264	999-99-9999		NA	1	✕
120-12-0085	999-99-9999		NA	1	✕
120-49-2078	999-99-9999		NA	1	✕
151-80-0426	999-99-9999		NA	1	✕
131-45-8031	999-99-9999		NA	1	✕
132-89-4131	999-99-9999		NA	1	✕
140-87-8252	999-99-9999		NA	1	✕

How we approached the "Right to erasure ('right to be forgotten')" and Indirect Identifiers

[Home](#)[Next Page](#)

- We use Hadoop as a staging area for interrogating data for privacy concerns as Hadoop removes the challenges that hide privacy concerns due to the complexity of data
- We use our intelligent catalog mechanisms to codify rules that define patterns of data that represent privacy concerns and schedule the interrogation of files and streams of data
- We use pattern detection to identify potential privacy concerns and use a process we call sequester, encrypt and secure, which encrypts the exposed data so that privacy information is not in harms way and sequester it into a highly protected environment
- We allow the creation of false positive lists so you do not have to revisit potential issues as new data enters your environment

Indirect Identifiers

[Home](#)[Next Page](#)

First we must establish joinable columns by Direct Identifiers (patterns in the data and business metadata)

Indirect Cross File Identifier Jobs									New Job
<input type="text" value="Search"/>									
Date/Time ▾	Server ▾	Port ▾	Path ▾	Delimiter ▾	Global Regex	Group Regex	Status	Action	
Sep 21, 2017 8:04:06 AM - Sep 21, 2017 8:07:37 AM	ns516032.ip-167-114-210.net	8020	/bigfiles			Click to view...		Share Delete	
Sep 14, 2017 6:14:03 AM - Sep 14, 2017 6:14:44 AM	ns516032.ip-167-114-210.net	8020	/binaryAndTextFiles			Click to view...		Share Delete	
Sep 8, 2017 8:16:10 AM - Sep 8, 2017 8:16:40 AM	ns516032.ip-167-114-210.net	8020	/FileFolder		Email	Click to view...		Share Delete	

Now we see what files are joinable by domain value (key value) email and what percentage of each files column has email

Identifier Summary						Run	Jobs
<input type="text" value="Search"/>							
Column ▴	Path ▾	Pattern ▾	Percentage ▾	Total Count ▾	Total Row ▾		
4	/binaryAndTextFiles/file1.csv	Email	75.00 %	9	12		
4	/binaryAndTextFiles/fileodf.ods	Email	52.94 %	9	17		
4	/binaryAndTextFiles/file2.csv	Email	76.92 %	10	13		
4	/binaryAndTextFiles/filedoc.docx	Email	75.00 %	9	12		
4	/binaryAndTextFiles/file3.csv	Email	75.00 %	9	12		
4	/binaryAndTextFiles/filexlsx.xlsx	Email	60.00 %	9	15		
4	/binaryAndTextFiles/filepdf	Email	52.63 %	10	19		
4	/binaryAndTextFiles/filehtml.html	Email	14.29 %	9	63		


Next
Page

Indirect Cross File Detail Jobs

Timestamp: Sep 14, 2017 6:14:03 AM

 Search

1 - 1 of 1

Identifier	Regex Count	Same File	Same File Count	Cross File	Action
dmorris6@narod.ru	SocialSecurityNumber=16,Gender=8,IPAddressValid=8	Yes	8	Yes	
jgriffin4@feedburner.com	SocialSecurityNumber=18,Gender=9,IPAddressValid=9	Yes	9	Yes	
jrussell5@who.int	SocialSecurityNumber=7,Gender=7,IPAddressValid=7	Yes	7	Yes	
ljenkins0@arstechnica.com	SocialSecurityNumber=2,Gender=1,IPAddressValid=1	No	0	Yes	
mdavis8@shutterfly.com	SocialSecurityNumber=6,Gender=6,IPAddressValid=6	Yes	6	Yes	
sh0016@gmail.com	SocialSecurityNumber=6,Gender=1,IPAddressValid=1	No	0	Yes	

Indirect Identifiers

[Home](#)[Next Page](#)

Now we selected the Indirect Identifier Basket of Patterns to find across multiple tables to Discover Indirect Identifiers or use this information to assist in the right to Erasure, “Right to be Forgotten”

Cross File Summary Results For Identifier : dmorris6@narod.ru						Consolidate	Encrypt
Select	Identifier	Row	Column	Path	Pattern		
<input type="checkbox"/>	dmorris6@narod.ru	8	9	/binaryAndTextFiles/file1.csv	Gender		
<input type="checkbox"/>	dmorris6@narod.ru	8	8	/binaryAndTextFiles/file1.csv	SocialSecurityNumber		
<input type="checkbox"/>	dmorris6@narod.ru	8	11	/binaryAndTextFiles/file1.csv	SocialSecurityNumber		
<input type="checkbox"/>	dmorris6@narod.ru	8	6	/binaryAndTextFiles/fileodf.ods	IPAddressValid		
<input type="checkbox"/>	dmorris6@narod.ru	8	9	/binaryAndTextFiles/fileodf.ods	Gender		
<input type="checkbox"/>	dmorris6@narod.ru	8	8	/binaryAndTextFiles/fileodf.ods	SocialSecurityNumber		
<input type="checkbox"/>	dmorris6@narod.ru	8	11	/binaryAndTextFiles/fileodf.ods	SocialSecurityNumber		
<input type="checkbox"/>	dmorris6@narod.ru	3	6	/binaryAndTextFiles/file2.csv	IPAddressValid		
<input type="checkbox"/>	dmorris6@narod.ru	3	9	/binaryAndTextFiles/file2.csv	Gender		
<input type="checkbox"/>	dmorris6@narod.ru	8	6	/binaryAndTextFiles/file1.csv	IPAddressValid		

1 - 10 of 32

[«](#) [1](#) [2](#) [3](#) [4](#) [»](#)

[10](#) [25](#) [50](#) [100](#)

Now we select the Key Domain for the file column and row we need to encrypt or sequester to not violate the GDPR Indirect Identifiers

Cross File Summary Results For Identifier : dmorris6@narod.ru						Consolidate	Encrypt
Select	Identifier	Row	Column	Path	Pattern		
<input checked="" type="checkbox"/>	dmorris6@narod.ru	8	9	/binaryAndTextFiles/file1.csv	Gender		
<input type="checkbox"/>	dmorris6@narod.ru	8	8	/binaryAndTextFiles/file1.csv	SocialSecurityNumber		
<input type="checkbox"/>	dmorris6@narod.ru	8	11	/binaryAndTextFiles/file1.csv	SocialSecurityNumber		
<input type="checkbox"/>	dmorris6@narod.ru	8	6	/binaryAndTextFiles/fileodf.ods	IPAddressValid		
<input checked="" type="checkbox"/>	dmorris6@narod.ru	8	9	/binaryAndTextFiles/fileodf.ods	Gender		
<input type="checkbox"/>	dmorris6@narod.ru	8	8	/binaryAndTextFiles/fileodf.ods	SocialSecurityNumber		
<input type="checkbox"/>	dmorris6@narod.ru	8	11	/binaryAndTextFiles/fileodf.ods	SocialSecurityNumber		
<input type="checkbox"/>	dmorris6@narod.ru	3	6	/binaryAndTextFiles/file2.csv	IPAddressValid		
<input type="checkbox"/>	dmorris6@narod.ru	3	9	/binaryAndTextFiles/file2.csv	Gender		
<input type="checkbox"/>	dmorris6@narod.ru	8	6	/binaryAndTextFiles/file1.csv	IPAddressValid		

1 - 10 of 32

[«](#) [1](#) [2](#) [3](#) [4](#) [»](#)

[10](#) [25](#) [50](#) [100](#)

Indirect Identifiers

[Home](#)[Next Page](#)

Here we can see we encrypted the proper row and column for this file for dmorris6@narod.ru.

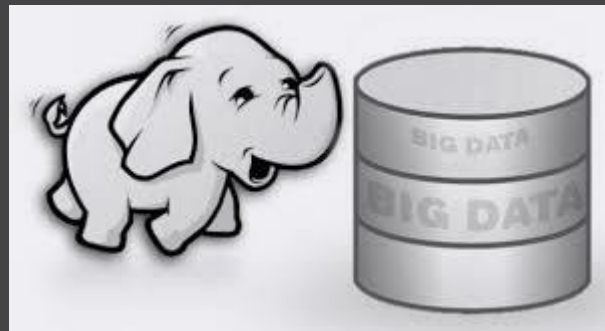
Now we select the Key Domain for the file column and row we need to encrypt or sequester to not violate the GDPR Indirect Identifiers

First Twenty Row From Latest Modify File Content:										
id	first_name	last_name	email	country	ip_address	phone_number	credit_card	gender	race	ssn
1	Lisa	Jenkins	ljenkins0@arstechnica.com			3-(077)695-6786			Indonesian	
2	James	Marshall	jmarshall1@bloglines.com				67098475265817698	M	Alaska Native	422-14-3562
3	Judith	Fowler	jfowler2@tiny.cc	Philippines		0-(751)958-0053	3546183479556523	F		114-17-2006
4	Michelle	Shaw			166.87.11.96	5-(206)112-9176	30596520921504	F	Cheyenne	164-20-4269
5	Jennifer	Griffin	jgriffin4@feedburner.com	Russia	199.243.4.20	4-(391)481-9509	3557104544425565	F	Cuban	531-20-0883
6	Janet	Russell	jrussell5@who.int	Russia	161.22.124.241	6-(378)138-6740		F	Sioux	343-56-3475
7	Debra	Morris	dmorris6@narod.ru	Ukraine	157.91.106.208		6385119534963459	iD81ZhJ3+YGH3iilUH9JwA==	Melanesian	999-99-1111
8	Christine	Murphy		South Africa	46.114.2.1	1-(167)727-4654	5578497981809253		Melanesian	932-95-8949

"Right to erasure ('right to be forgotten ')"

GDPR requires companies to maintain verification that a customer has given consent for you to use their Personal Information and also requires you to provide a graphical interface for them to later request erasure of that information. Consent and Erasure can be unlimited or selective in nature; meaning a customer may wish to receive email coupons but not allow use of their home phone or address.

Consent – Request for Erasure
Graphical interface and API's



Consent Storage
Right to be Forgotten Storage
Third Party Consent Storage



"Right to erasure ('right to be forgotten')"

- (1) The data subject shall have the right to obtain from the controller the erasure of personal data concerning him or her without undue delay and the controller shall have the obligation to erase personal data without undue delay where one of the following grounds applies:
 - (a) the personal data are no longer necessary in relation to the purposes for which they were collected or otherwise processed;
 - (b) the data subject withdraws consent on which the processing is based according to point (a) of [Article 6\(1\)](#), or point (a) of [Article 9\(2\)](#), and where there is no other legal ground for the processing;
 - (c) the data subject objects to the processing pursuant to [Article 21\(1\)](#) and there are no overriding legitimate grounds for the processing, or the data subject objects to the processing pursuant to [Article 21\(2\)](#);
 - (d) the personal data have been unlawfully processed;
 - (e) the personal data have to be erased for compliance with a legal obligation in Union or Member State law to which the controller is subject;
 - (f) the personal data have been collected in relation to the offer of information society services referred to in [Article 8\(1\)](#).



"Right to erasure ('right to be forgotten')"

Cont.

(2) Where the controller has made the personal data public and is obliged pursuant to paragraph 1 to erase the personal data, the controller, taking account of available technology and the cost of implementation, shall take reasonable steps, including technical measures, to inform controllers which are processing the personal data that the data subject has requested the erasure by such controllers of any links to, or copy or replication of, those personal data.

(3) Paragraphs 1 and 2 shall not apply to the extent that processing is necessary:

- (a) for exercising the right of freedom of expression and information;
- (b) for compliance with a legal obligation which requires processing by Union or Member State law to which the controller is subject or for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller;
- (c) for reasons of public interest in the area of public health in accordance with points (h) and (i) of [Article 9](#)(2) as well as [Article 9](#)(3);
- (d) for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with [Article 89](#)(1) in so far as the right referred to in paragraph 1 is likely to render impossible or seriously impair the achievement of the objectives of that processing; or
- (e) for the establishment, exercise or defence of legal claims.

Calculating Benchmarks

[Home](#)[Next
Page](#)

Things to consider when calculating Benchmarks and Latency:

Benchmarks:

1. Number of Files
2. Average number of columns
3. Folder Configurations
4. Type of Data:
 1. Structured
 2. Unstructured
 3. Binary
 4. Email
 5. Others
5. The various processes to be run:
 1. Quick Column Classification with Intelligent Catalog and Metadata
 2. Adding Column Headers from the Intelligent Catalog
 3. Complete Discovery run of All Data
 4. Review of results
 5. Deciphering through the graphical interface what are false positives versus real risks and violations
 6. Deciding if a complete file should be sequestered or if specific columns / rows should be encrypted
 7. Search for Indirect Identifiers contained in numerous files of various formats and structures by logically joining them by using other Direct Identifiers as Keys.
 1. Review results that identify which files, when joined together, contain enough indirect identifiers to positively identify an individual or small block of individuals

Calculating Benchmarks

[Home](#)

Things to consider when calculating Benchmarks and Latency,
Continued:

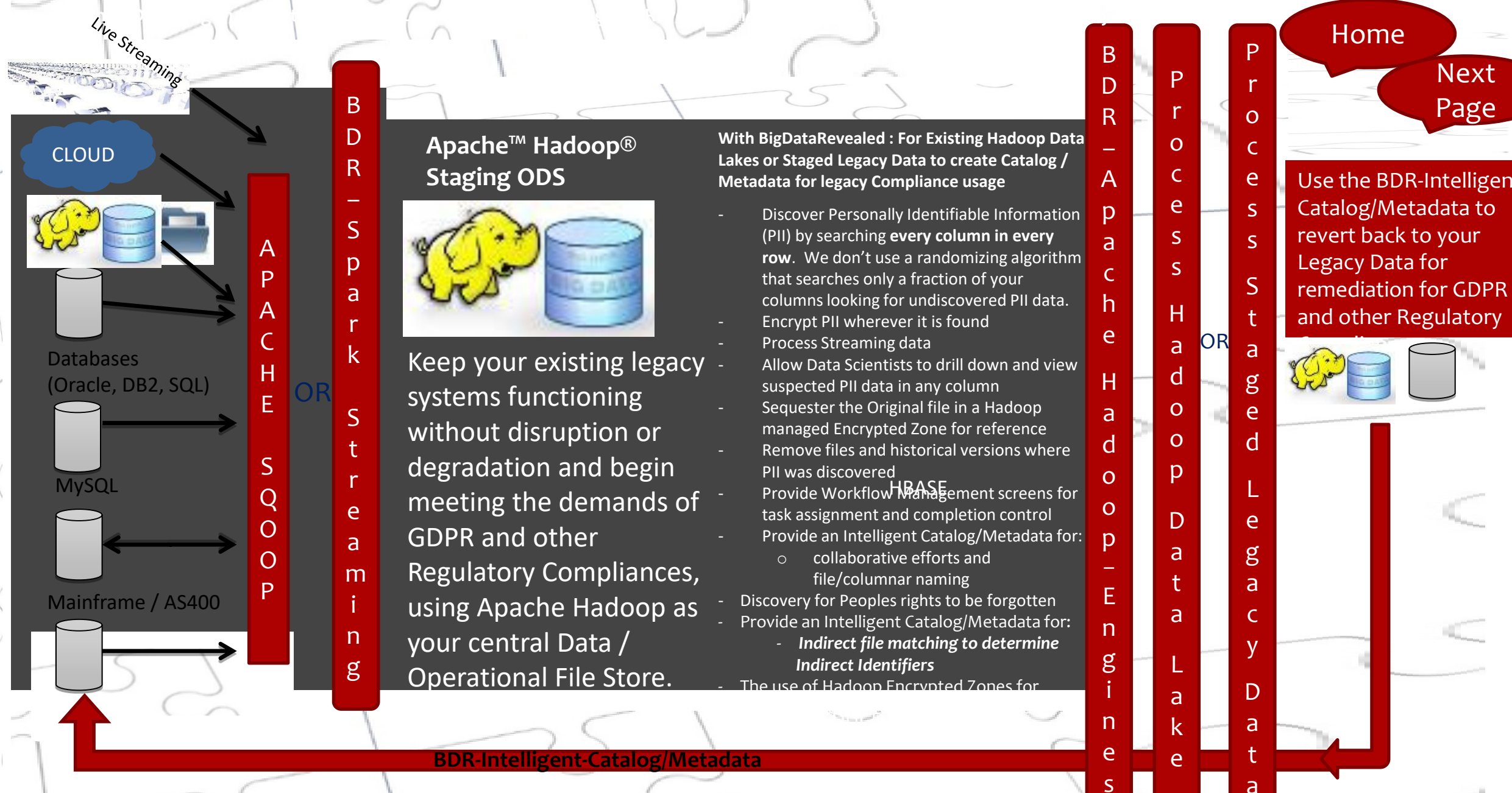
2. Determine which columns or rows needs to be encrypted, or which files need to be sequestered, even if the fields reside on files of unlike structure or type.
 3. Use the list of individuals exercising their 'Right to be Forgotten' to further select fields to encrypt or files to sequester.
-
8. Volume test using a variety of files, file types, file sizes / rows, data types, with server clusters containing a various number of nodes. Document the time it took these processes to run in a repeatable manner closest to zero latency as possible.

Our benchmarking for jobs that performed discovery and remediation by encryption on one Node, with 50 million rows has executed in under two minute for most processes. On a 4 node box 1 billion rows was processed in just a few hours. Indirect identifiers will depend on the number of files and columns and total permutations to get 100% of the results. We found 10 million rows with 3-5 files on one node still runs in under an hour.

Our product is 100% Spark / Java 8 jobs running as part of the Hadoop framework and eco-system so that data Never leaves the Hadoop ecosystem. We scale 100% in parallel with the nodes and power of your Hadoop platform.

Also consider running live streaming data and performing discovery and remediation on the fly so that you are comfortable data will be secured before it is even loaded into HDFS or Hbase.

BigDataRevealed Architecture For GDPR / Legacy - for All Regulatory Compliances, Powered by Apache™ Hadoop®



Productionalizing Methodology

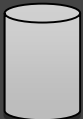
CLOUD



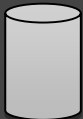
Live Streams (Spark)



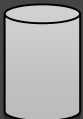
Office Documents



Databases:
(Oracle, DB2,
SQL Server,
Teradata, Flat File) ...



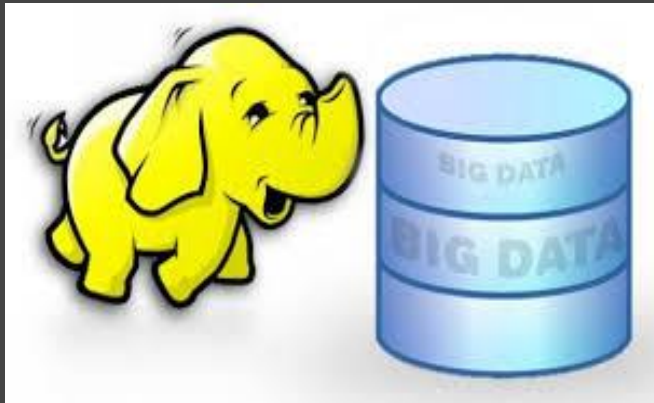
MySQL



Mainframe / AS400

Quick Pattern Business
Classifications with Intelligent
Catalog / Metadata

Audit / Assessment Centralized Repository
Apache Hadoop Recommended



BigDataRevealed
Processes
&
Productionizing

Home

Quick
Pattern &
Intelligent
Catalog /
Metadata

- Simply Run Quick Classification on Folders
- View results, Export Results, Share Results

Review of
Quick Pattern
Analysis of
Metadata

- Based on Results Prepare Plan for remediation
- Know how many files / columns need processing

Calculate
Benchmarks
and best
latency

- Calculate total files / columns and rows to be encrypted
- Calculate baseline latency and timing for production

Setup live
streaming
processes

- Complete the above processing for live streams

Setup and
Run
Complete
Pattern PII
processes

- Based on analysis select compliance algorithms and files
- Run Discovery, remediation and schedule best latency

Remediation
Encryption
Sequester

- Encryption of columns/rows, complete columns or files
- Sequester binary and other files as needed in Encrypted Zones

Right of
Erasure / to be
Forgotten

- Select keys to join files to complete Indirect Identifier detection.
- Join all file types, including binary files, by selected keys and process Cross File Indirect Identifier discovery.
- Encrypt file/column/row or sequester files to be compliant

Push new data or complete data
over at the shortest latency possible

The roadmap to compliance

BDR is the engine that houses the intelligent catalog and the GDPR enablement kit for discovery and protection of PII data as mandated by GDPR. It also publishes a searchable secured catalog, which allows for compliance with tight deadlines provisioned to respond to consumers wishing to be forgotten.

We recommend starting with a Privacy Self Assessment which we will help analyze and benchmark against organizations similar to yours during an intensive 3 day session. This \$5K cost of this session will be waived when engaging to install BigDataRevealed.

Our current list price for the Privacy Assurance solution initially geared for GDPR is an annual recurring cost of \$20K for the master node and \$1K for any additional nodes. We are discounting the investment for early adopters who install and provide us with testimonials demonstrating how BigDataRevealed helped in their journey to privacy assurance.

To encourage compliance and make a examples of non-compliant companies the EU has promised fines as stiff as 4% of your prior year's revenues. We believe a sound GDPR solution armed with a holistic approach to privacy assurance is the only prudent means to demonstrate to the EU that you have your compliance house in order.

We encourage you to start the journey with us by clicking on the privacy assurance self assessment

[CLICK HERE →](#) [Privacy Assurance Calculator Excel](#)

Home