Discovery-in-Depth

Results, analysis, and key takeaways from BigID's
discovery-in-depth for enterprise data environments

Discovery-in-Depth Overview

Discovery-in-Depth: How it Works

BigID scans data across structured, unstructured, and semi-structured stores - in cloud, on-prem, and hybrid environments.

Leveraging ML-based discovery-in-depth, BigID identifies, inventories, and classifies sensitive, personal, and regulated data; uncovers duplicate data; and highlights vulnerable data.

How? With 4Cs: Classification, Catalog, Cluster Analysis, and Correlation.

Benefits:

✓ Full visibility of all data, everywhere

✓ Automatically discover sensitive, personal, and regulated data

✓ ML-driven data insight & intelligence

✓ Unified data inventory for context, coverage, and transparency

✓ Build a foundation from which to take action

Supported Data

✓ Unstructured

✓ Structured

✓ Semi-structured

✓ At-rest

✓ In-motion

see page 9 for sample coverage

✓ On-prem

✓ Cloud

✓ Hybrid Environments

DISCOVERY DOWNLOAD

Sample Findings at-a-glance

6.12M personal records identified

847,332 email addresses discovered

12,839 PCI records tagged

12% all data open to everyone

19% all data was duplicate or redundant

84% of personal data discovered falls under CCPA, GDPR, and GLBA regulations

28 OOB Classifiers identified 629,391 sensitive data attributes across MongoDB, Office365, and AWS S3.

NER identified 526 financial statements, 14 invoices, and 4 M&A documents.

BigID applied patented ML-based discovery-in-depth techniques to automatically identify, classify, and label regulated data to enable ongoing compliance with finance industry regulations on a sample of data stores: including MongoDB, Office 365, and AWS S3.

Native classifiers identified that 47% of ▮▮▮▮▮▮▮▮▮▮ data in their AWS S3 environment (based on a sample) was personal information (PI).

The data security group was able to preview a unified inventory where they can manage both structured and unstructured data all in one place for an upcoming file analysis and security framework initiative.

Classification Spotlight

Discover personally identifiable information & sensitive data with BigID pre-configured classifiers - across structured, semi-structured and unstructured data sources.

This technique leverages traditional pattern matching & regular expressions: it's vulnerable to a high volume of false-positives and inaccuracies, and requires correlation and cataloging layered in to gain validation and trust in the data.

Benefits:

✓ Report on regulated data

✓ Tag and label PII by type or regulation

✓ OOB classifiers

✓ Create custom classifiers

sample findings

Classification	Count
Credit Card	22,393
Email	138,029
Explicit passwords	4,121
IPv4	19,200
Product keys	37
Gender	44,813
Social Security Numbers	67,921

BigID

NER Document Spotlight

BigID applies Named Entity Recognition (NER) to automatically identify specific types of documents that typically contain sensitive and regulated data. BigID automatically recognizes document types - based on what's inside - in order to discover and classify sensitive data in an additional layer of data intelligence.

Sample document types include financial statements, invoices, legal agreements, RFPs, criminal histories, and more.

Benefits:

✓ Automatically identify sensitive documents

✓ Prioritize data management based on sensitivity

sample findings

| Document Type	Count
Financial Statement	31,399
Invoice	44
M&A Document	12
Federal Income Tax Return	7,211
NDA	912

Correlation Spotlight

Correlation finds more data, more accurately - without having to know where or what it is. BigID can look for a few attributes across the structured data sources - and discover both related data as well as data elements that cannot be discovered with metadata scanners or by regular expression.

Benefits:

✓ Automatically find dark data

✓ Build identity profiles for DSAR workflows

✓ Discover PI and sensitive data without patterns

✓ Identify relationships between data
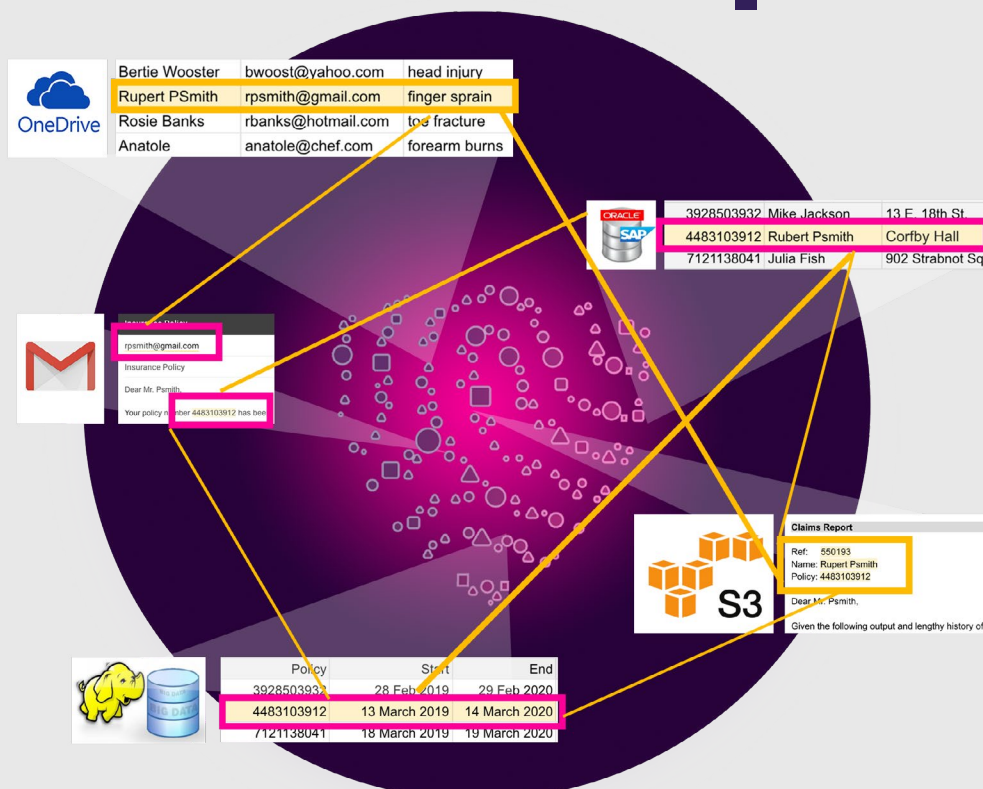
sample findings

Duplicate Data Spotlight

Cluster analysis automatically identifies duplicate and redundant data, which poses hurdles for data quality, increases risk, and amplifies noise over insight for data analysis.

BigID automatically clusters similar and duplicate data for faster time to value with 93% accuracy.

Benefits:

✓ Clean up duplicate and redundant data

✓ Reduce risk on vulnerable data

✓ Improve data quality
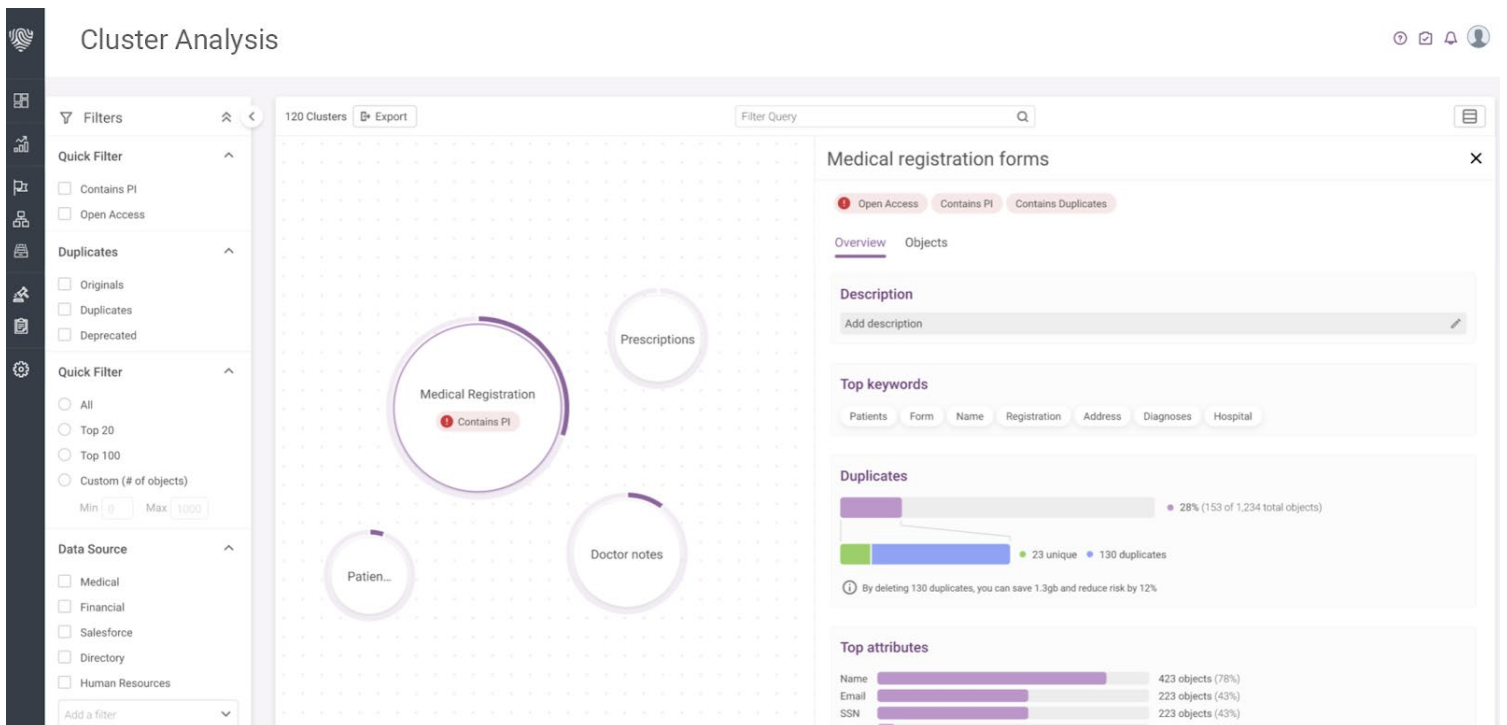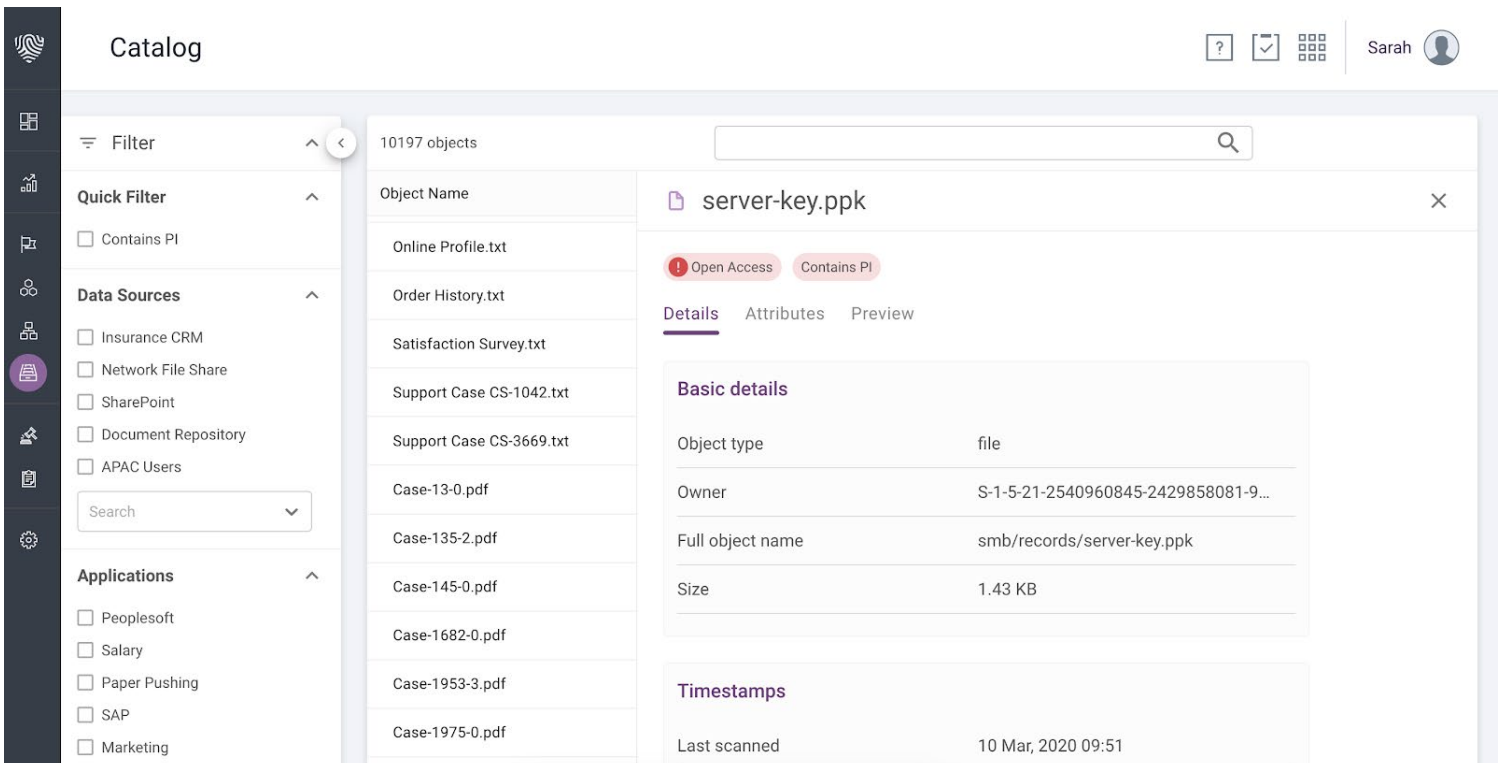
sample findings

Catalog Spotlight

BigID's catalog view gives more context and more data intelligence: see personal & sensitive data in context with technical, business, and operational metadata.

Enrich findings and get a unified view from across all of your data sources - all in one place.

Benefits:

✓ Enrich data catalog findings

✓ Data quality scoring

✓ Leverage active metadata

✓ Search and sort with ease

sample view

Catalog

Sarah

Filter		10197 objects	
Quick Filter		Object Name	

Quick Filter
☐ Contains PI

Data Sources
☐ Insurance CRM
☐ Network File Share
☐ SharePoint
☐ Document Repository
☐ APAC Users

Search

Applications
☐ Peoplesoft
☐ Salary
☐ Paper Pushing
☐ SAP
☐ Marketing

Object Name
Online Profile.txt
Order History.txt
Satisfaction Survey.txt
Support Case CS-1042.txt
Support Case CS-3669.txt
Case-13-0.pdf
Case-135-2.pdf
Case-145-0.pdf
Case-1682-0.pdf
Case-1953-3.pdf
Case-1975-0.pdf

server-key.ppk

⚠ Open Access   Contains PI

Details   Attributes   Preview

Basic details

Object type	file
Owner	S-1-5-21-2540960845-2429858081-9...
Full object name	smb/records/server-key.ppk
Size	1.43 KB

Timestamps

Last scanned	10 Mar, 2020 09:51

Coverage Spotlight

BigID's data coverage leverages agentless native BigID connectors across your data center & the cloud, supporting data-at-rest and data-in-motion.

We cover more data types across more data sources: including structured databases, unstructured file shares, big data & NoSQL, messaging/communications, email, middleware, cloud SaaS, cloud IaaS, mainframe, applications, and data in motion.

Benefits:

✓ All data, everywhere

✓ At rest & in motion

✓ On-prem, in the cloud, & hybrid environments

✓ Single pane of glass

sample coverage

Modern Architecture

BigID runs on modern architecture and is micro-service by design: built around a Docker and Kubernetes core that can easily be deployed in the data center, cloud or both.

Built to scale, BigID is designed to manage large volumes of data across an organization's data landscape. With BigID, companies can deploy data discovery in a data center or cloud for almost any data.

Benefits:

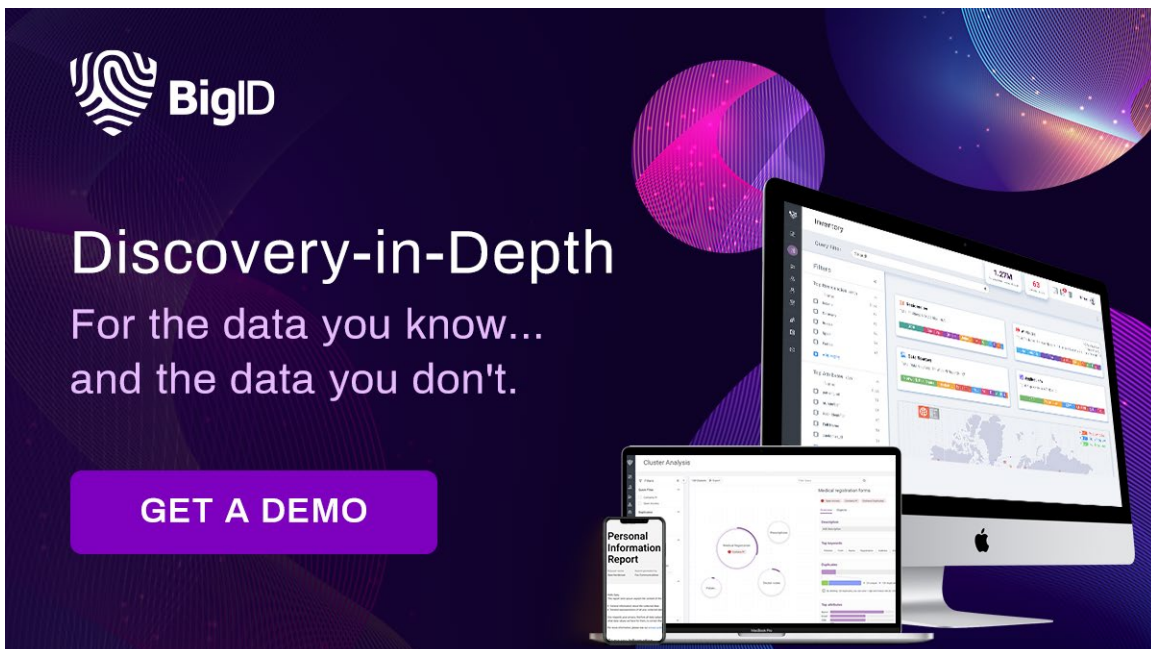✓ Multiple deployment options

✓ Agentless

✓ Built to scale

✓ Easy orchestration

sample deployment
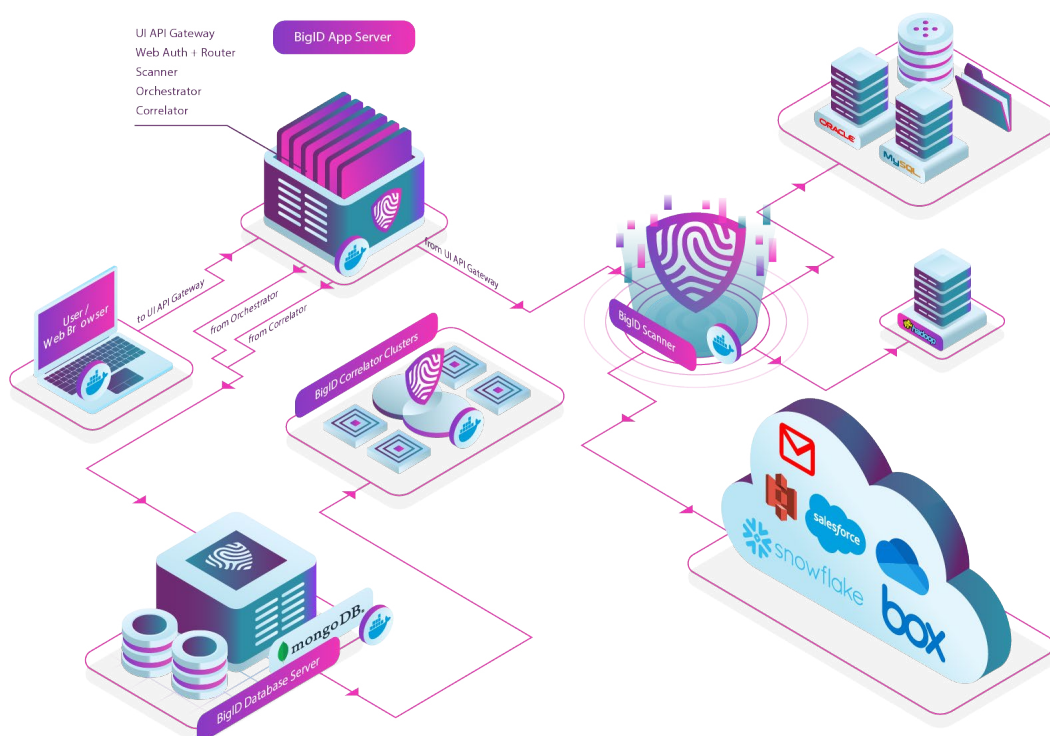
Key Takeaways

▶ **Automatically Discover, inventory, and map** all types of personal and sensitive data

▶ Leverage **advanced ML techniques** including NER and deep learning

▶ Go beyond compliance checklists with an end-to-end **framework for privacy & risk**

▶ Fulfill **governance and catalog initiatives** with a **unified inventory** aligning personal, sensitive, and metadata discovery

▶ Discover **dark data** and **protect crown jewels** with intelligent insight and remediation

▶ **Reduce risk** with 360 visibility

Discovery-in-Depth

For the data you know... and the data you don't.

GET A DEMO

BigID Data Intelligence Platform

How It Works

Unmatched Data Coverage + Discovery-in-Depth + Insights + Actions

Data Privacy Apps

- PI / PII Inventory
- DSAR Fulfillment Automation
- Data Processing & Sharing
- Consent Governance
- Privacy Portal

Data Protection Apps

- Remediation
- Data Access Intelligence
- Labelling
- Breach Data
- Data Risk

Data Perspective Apps

- Data Quality
- Data Retention
- Stewardship Self-service
- Advanced Profiling
- Data Lineage

Data Discovery Foundation

Discover, identify, and classify all of your data, everywhere -
with BigID's patented Discovery-in-Depth technology

| Catalog | Classify | Cluster Analysis | Correlate |

Unstructured · Structured · Semi-structured · Big Data · NoSQL · Data Lakes ·
Data Warehouses · OCR Applications · SAP · Email · Messaging · Mainframe ·
Data Pipelines · IaaS · SaaS · APIs · Network