



Know Your Data.

Privacy • Security • Governance

Privacy by Design

The Advent of Privacy Engineering

New York - Tel Aviv - London - Paris - Zurich - Singapore - Sao Paulo - Bangalore

Agenda

- Introductions
- Privacy by Design and Privacy Engineering Intro
- BigID & IBM Cloud Pak for Data (CP4D) Status/Update
- BigID Ecosystem/Alliances



-
- BigID Solution Overview
 - BigID Demonstration
 - Q&A

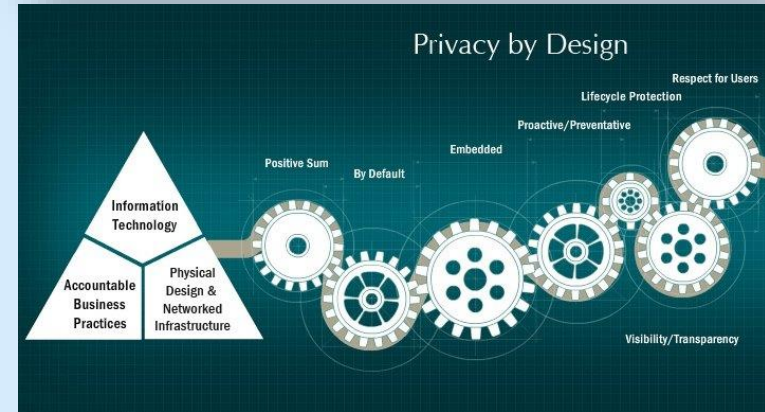
Privacy by Design (PbD) for Enhanced Customer Trust

- Data and its Privacy is a Strategic Priority for success in the Digital Age.
- Holistic Privacy Strategies deliver Benefits beyond Regulatory Compliance (GDPR/CCPA)
- Privacy by Design & Default
 - Developer Friendly - Privacy Shift Left
 - Advent of Privacy Engineers (Technologists)
 - Confident Firms apply Data Privacy Best Practices and strong Technology Controls
- Establishing Privacy Governance
 - Delivering a Holistic Data Protection Approach is key to Success.
 - Transform Data Risk into Business Value
 - Data Intelligence & Identity Data Awareness as a "Force Multiplier" to existing Security Frameworks.
- Enterprise Data Privacy & Security Programs need to have Global Scope & Scale

Data Privacy Is The New Strategic Priority

FORRESTER®

Enhanced Customer Trust Is The No. 1 Advantage Decision Makers Expect From Adopting A Holistic Data Privacy Approach



Key Findings



Privacy is a strategic priority for today's organizations, but few have complete confidence in their ongoing ability to keep up with emerging data privacy regulations.



Siloed data privacy projects are becoming increasingly inefficient and unstable. Sustained compliance needs an operationalized data privacy program that is holistic, scalable, and global in scope.



Organizations with well-rounded strategies that commit to data privacy for more than just compliance stand to realize a number of benefits, including enhanced customer trust.

The Advent of Privacy Engineering

Standardizing Privacy By Design (PbD) And Privacy Engineering

- Move over Attorneys, CPOs & DPOs – Make room for Privacy Engineers
- PbD – A Set of Principles to Embed Privacy & Data Protection into Products & Services
- Privacy Engineers are the “Implementers” of PbD
- Data Scientists & Analysts ability to select “PET”
- Industry Privacy Standards ISO/PC 317 PbD / NIST
- Gain & Maintain Trust – High Demand for Privacy Savvy Technologists and Privacy Engineers

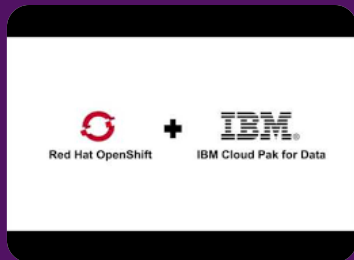


To understand the core capabilities of privacy engineering for product development, we can take a look at how the core capabilities that **Carnegie Mellon University** prepares its students to apply upon completion of its **Masters of Science in Information Technology — Privacy Engineering Program**

- 1) Design cutting-edge products and services that leverage big data while preserving privacy;
- 2) Propose and evaluate solutions to mitigate privacy risks;
- 3) Understand how privacy-enhancing technologies can be used to reduce privacy risks;
- 4) Use techniques to aggregate and de-identify data, and understand the limits of de-identification;
- 5) Understand current privacy regulatory and self-regulatory frameworks;
- 6) Understand current technology-related privacy issues;
- 7) Conduct privacy-related risk assessments and compliance reviews, respond to incidents, and integrate privacy into the software engineering lifecycle phases;
- 8) Conduct basic usability evaluations to assess the usability and user acceptance of privacy-related features and processes; and
- 9) Serve as an effective privacy subject-matter expert, working with interdisciplinary teams.

BigID Profile – IBM CP4D Playground

A Modern Privacy Aware Data Discovery and Intelligence Platform built on Containers & Kubernetes



BigID A Modern Privacy Aware Data Discovery and Intelligence Platform [Deploy](#)

Inventory 1.33M 63

Query Filter: Search

Filters

- Source: 16
- Attributes: 16
- Data Sources: 16
- Applications: 16

Residences 10 Residences 17 Residences 20 Residences 20

Attributes 10 Attributes 17 Attributes 20 Attributes 20

Data Sources 10 Data Sources 17 Data Sources 20 Data Sources 20

Applications 10 Applications 17 Applications 20 Applications 20

Map

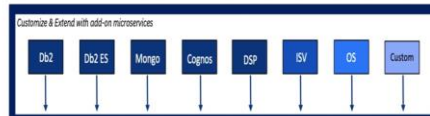
About the developer:

BigID

Built for Multi-cloud

Avoid vendor lock-in & get started on your cloud journey today

PICK YOUR ADD-ON Containerized Services



DATA & AI PLATFORM #1 Ranked by Forrester



KUBERNETES BASED Containerized, easy to manage



PICK YOUR CLOUD Private or Public



Modern Enterprise Architecture

Scalable, Flexible & Secure Deployments



RBAC
with Scoping



Credential
Externalization



Any Data /
Any Language



API-First
SDK for Design for Privacy



Containers with
Kubernetes



Pre-built 3rd Party
Orchestrations



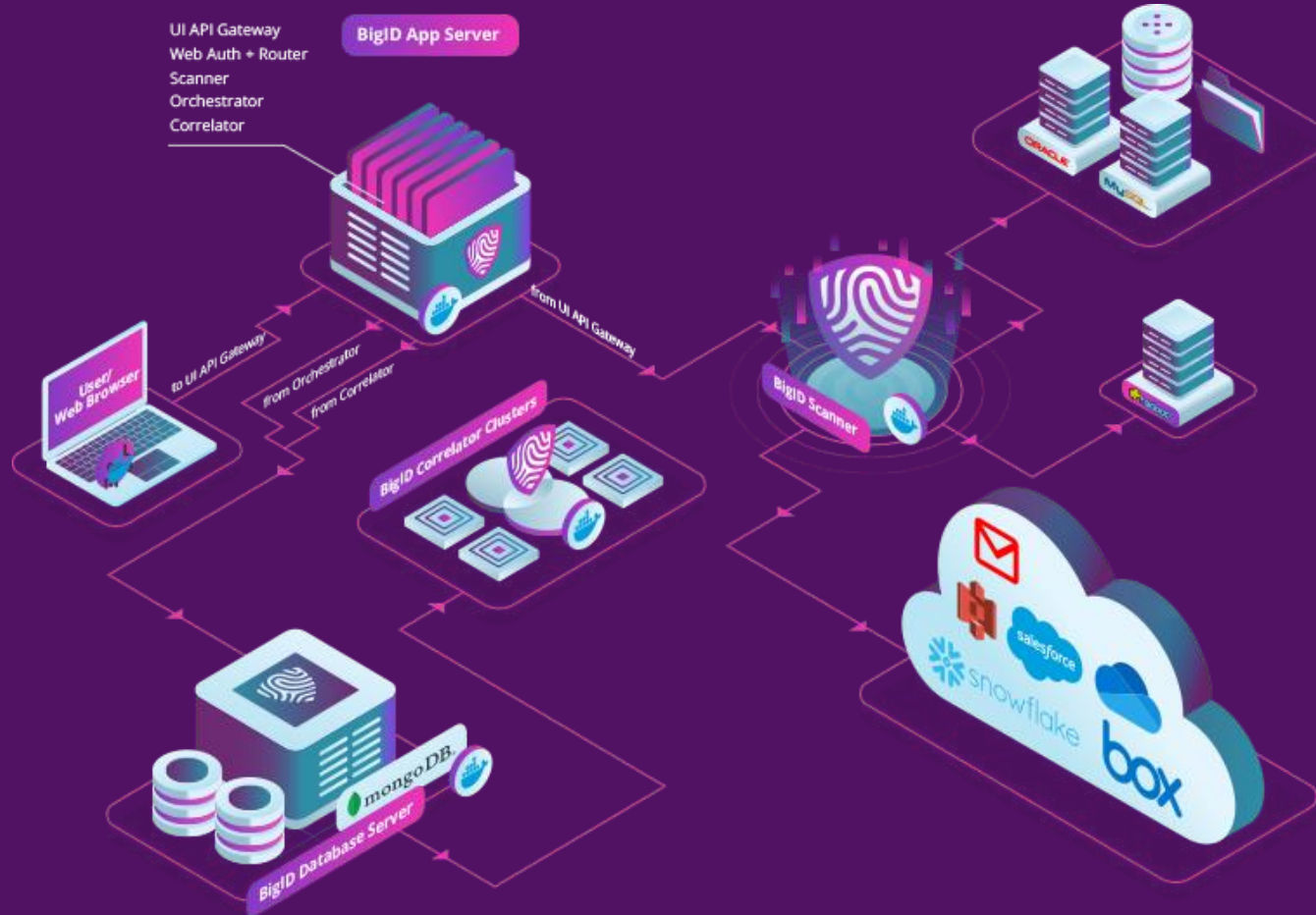
No Data
Copying



On-prem or
Multi Cloud Deployment



Modern Container & Kubernetes Architecture



BigID Strategic Alliances & ISV Partners



Leading Cloud Platforms and BigID Vendor Alliances



Enterprise Applications Cloud Migrations & Born on the Cloud

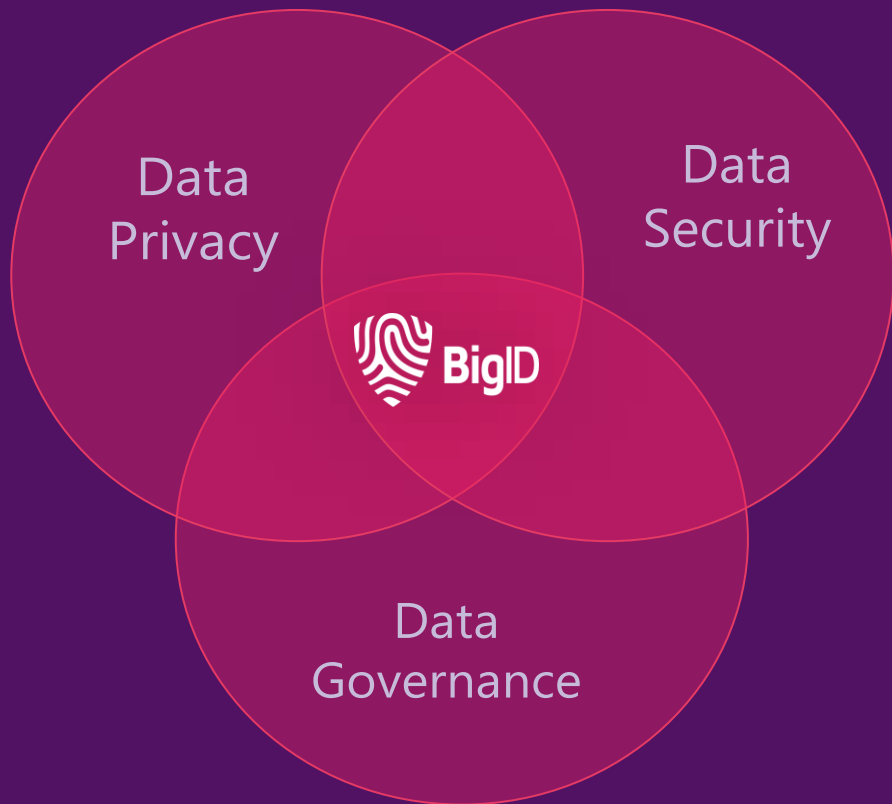


ISV Partners for E2E Solutions



Delivering Holistic Data Intelligence Solutions

Transform Data Risk into Business Value – Discovery to Insight



Privacy Requires Rethink of Data Discovery

Regular Expressions (RegEx) and Metadata Isn't Enough



Find PII and
Contextual PI

Classify, Catalog &
Correlate by Person



Automate Data Privacy
Security & Governance

- GDPR, CCPA/CPRA, HIPAA etc

Deeper Data
Intelligence to
include ML/AI



Privacy Aware Data Discovery & Intelligence Platform

Application Framework

Data Privacy Apps

- PI / PII Inventory
- DSAR Fulfillment Automation
- Data Processing & Sharing
- Consent Governance
- **Privacy Portal (Self Service)**

Data Protection Apps

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

Data Perspective Apps

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

Data Discovery Foundation – Discovery in Depth

Catalog

Classify

Cluster Analysis

Correlate

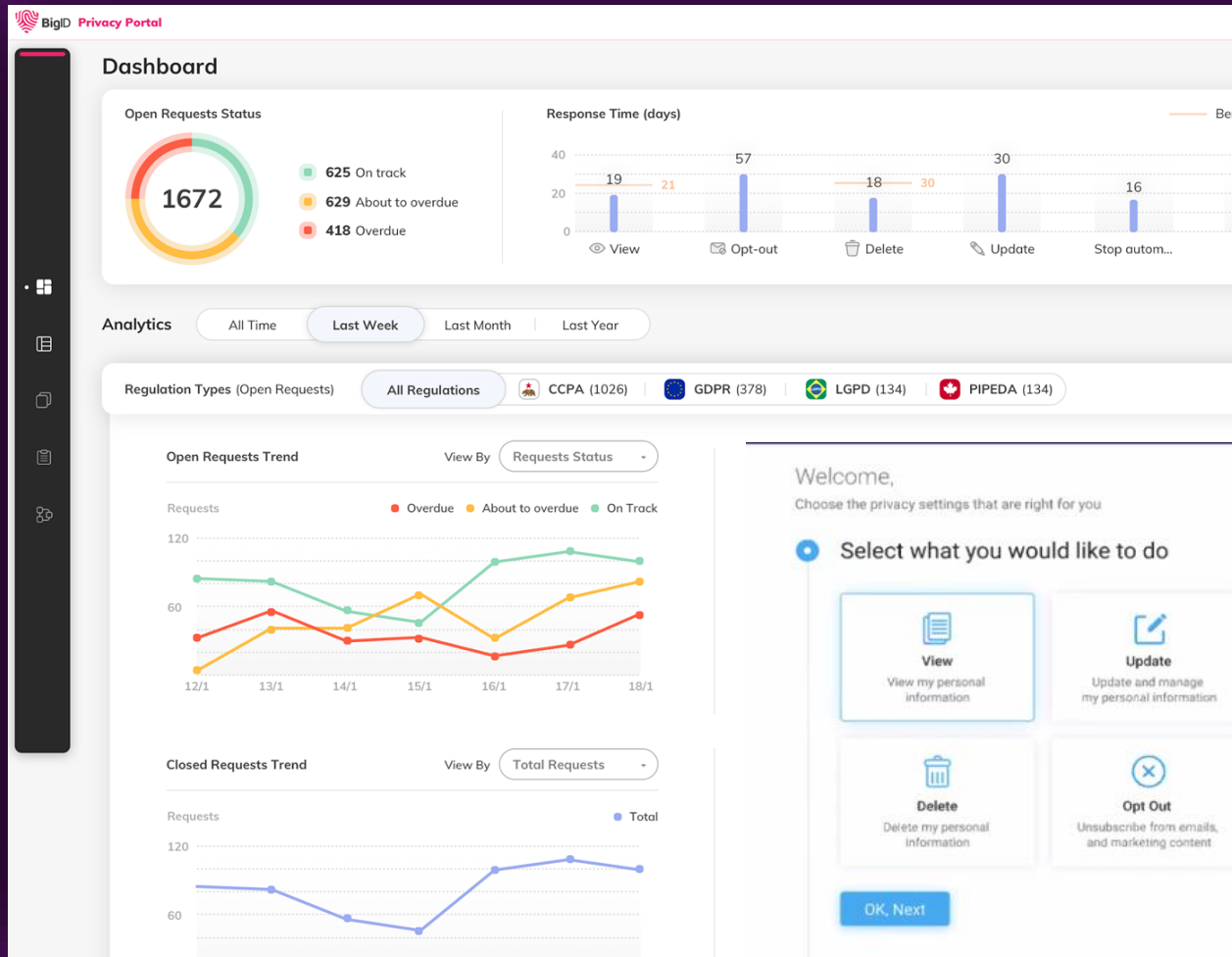
Unstructured · Structured · Semi-structured · Big Data · NoSQL · Data Lakes · Data Warehouses · OCR
Applications · SAP · SFDC · Email · Messaging · Mainframe · Data Pipelines · IaaS · SaaS · APIs · Network



USE CASE

Privacy Portal DSAR Request Management

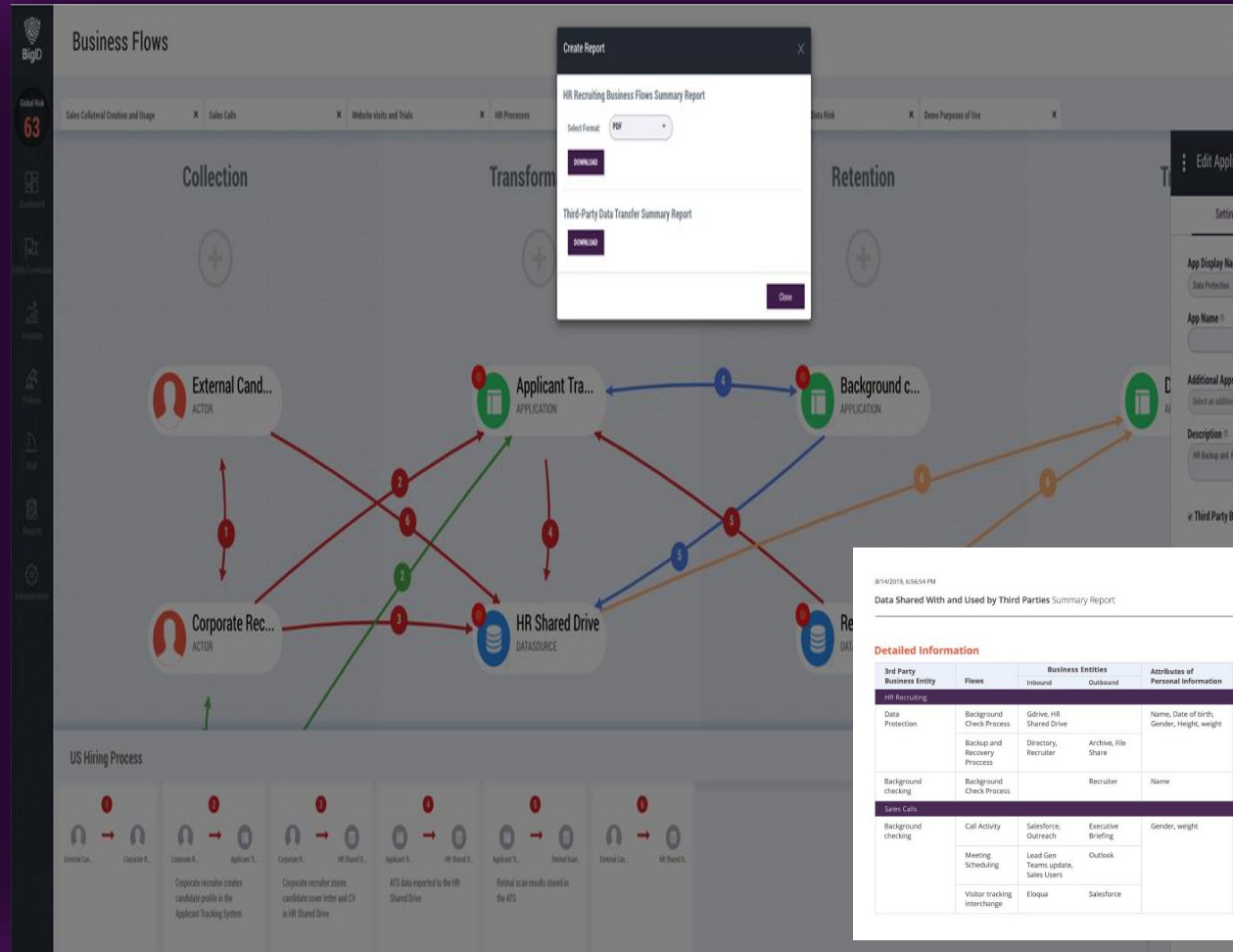
- **End-to-end Workflow**
Manage DSAR request lifecycle from intake to fulfillment
- **Automate Processes**
Automate request approval, task delegation, deletion workflows
- **Track progress**
Manage and monitor request status
- **Embedded** into other 3rd Party solutions, workflows and UIs



USE CASE

Automate 3rd Party CCPA & GDPR Sharing

- Automate How Companies Build & Maintain Data Flow Maps
- Record Data Processing Activity & Track Data Usage
- Track CCPA, GDPR External Data Sharing.
- Scan Data in Motion for PI and trigger protection actions



USE CASE

OCR

Scan images and scanned documents using OCR to detect personal identifiers & sensitive data

Include Scanned data in DSAR

Classifications And Findings

103 Classifiers

Classifier Name	
LineageGuid	163,671
Email	44,688
IPv4	16,672
US SSN	2,486
email #2	751
NER::Country/City	720
NER::Full Name	412
NER::Phone	88
GeographicData	13
Product Keys	1
Explicit Password	1
Textual PPK Private Key	1

LineageGuid - 163,671

Data Sources

Insurance CRM

?

✓ 14

Grid icon

Responsible

Data



Applied

USE CASE

BigID HyperScan

Intelligent scan of
unstructured data sources

- Patent-pending machine learning based algorithm
- Predict if a file or data set contains sensitive data based on metadata only
- Scan only sensitive files
- Save 80-90% of scan time!




< Administration / New Scan Profile

Name


Description

Scan Type

Hyper Scan



Owner Email (1)



[+ Add another owner](#)

☐ Schedule

Every:

☒ All enabled data sources ☐ Select data sources

1

Sample Scan
Smart sampling of 5-10% of files. Scanning process includes both content and metadata

2

Metadata Scan
Run Metadata scan on all files (10X faster than full scan).

3

Calculate Model
Create a predictive model based on metadata features. Model features include owner, name, path, size etc.

4

Apply Predictions
Report findings in Catalog. For each file, assign probability of finding the sensitive attribute.

COMING SOON

BigID HyperScan (Predictive)

Sample
Scan



Metadata
Scan



Generate
Predictive Model



Apply
Predictions

- Scan 5% of the data with 95% accuracy
- Configurable Recall (false negatives) and Precision (false positives)
- Option to automatically scan files based on confidence level

	Recall	Precision	Saving
Customer 1 - Box	0.95	0.53 (1*)	90% (50%*)
Customer 2 - S3	0.95	0.96	90%
Customer 3 - HDFS	0.95	0.96	90%



* Running an additional scan only on the files with predicted findings brings precision to 100%

BigID Developer Tools

Add Connectors, Apps, Integrations, Pipelines and More

APIs

BigID customers & partners can access, experiment and explore BigID REST APIs

SDKs

BigID customers & partners can build and share BigID connectors

APPs

BigID customers & partners can build and share BigID enhancements and integrations

The screenshot shows the BigID Developer Tools website. The header includes the BigIDExchange logo and navigation links: Community, Download, Meet, Develop, Forums, Blog, Learn, and Watch. A search icon and a 'GETTING STARTED' button are also present. The main content area features a large banner for 'Data Sources Discoverer' with a globe background and a magnifying glass. Below the banner, there are sections for 'Meet' (March 2020 Meetup - Community Launch), 'Develop' (Best Practice for Content Development and BigID Operationalization Guide), 'Forums' (Forum Rules, BigID Release 2019-10-13 is out!), and 'Events' (BigID at CyberTech TLV, March 2020 Meetup, BigID 2020 SKO).

BigIDExchange | Community | Download | Meet | Develop | Forums | Blog | Learn | Watch

GETTING STARTED

Download

See all >

Data Sources Discoverer
Discover and import data sources using NMAP/from Service-Now/from AD
★★★★★

Meet

See all >

March 2020 Meetup - Community Launch
Register Now!

Develop

See all >

Best Practice

- Best Practice for Content Development
- BigID Operationalization Guide

Forums

Forum Rules
BigID Release 2019-10-13 is out!
Installation fails
More

Events

- Jan 30 BigID at CyberTech TLV Tel Aviv
- Mar 03 March 2020 Meetup <https://zoom.us/j/12345678>
- BigID 2020 SKO

BigID.io

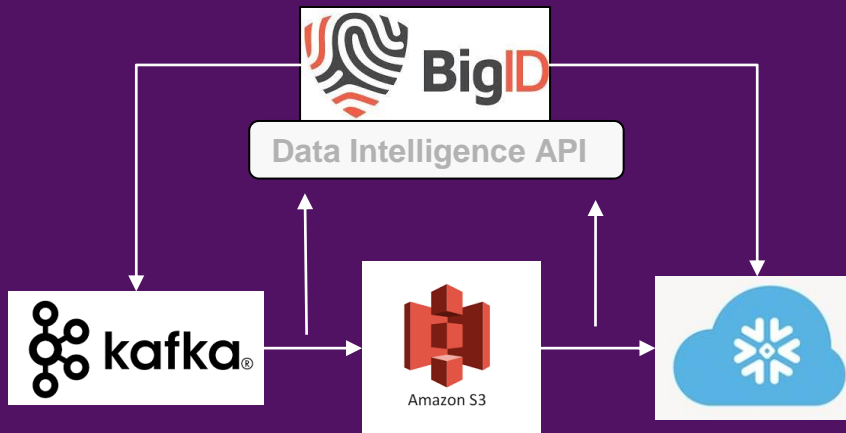
APIs for developers to build privacy and data protection into data pipelines & applications

Build data pipelines with privacy by design (CTO)

Gain data insights about the data in the pipeline (CTO)

Build the data catalog without friction (CDO)

Monitor and Control Data Sharing (CPO, CISO, CDO)



Scan APIs for Developers

- **Shift Left**
Simplify developer support
- **Cloud-native**
Simplify catalog, classify or correlate in cloud native services
- **Scan Payloads Async**
Scan payloads in FTP etc
- **Scan Pipelines & APIs**
Scan Kafka, APIs, Kinesis etc

 SUGGEST EDITS

REQUEST: Formatted payload containing data to scan.

RESPONSE: Return code that data is valid (202 Accepted) or invalid (error code).

POST

<https://BigID/api/v1/scan-api/analyze>



🔗 Try It

BODY PARAMS

channel* string

Name for the payload, maps to Data Object in inventory.

Example: Incoming Leads

payload* string

Data to scan.

Example: {user: simon, email: simon@example.com}

cURL Node Ruby JavaScript Python

```
curl --request POST \  
--url https://bigid/api/v1/scan-api/analyze \  
--header 'accept: application/json' \  
--header 'authorization: Basic OmJpZ2lkZG9yOjEwMTU0MzY1Njc4OTQ=\  
\ --header 'content-type: application/json'
```

Try the API to see Results

RESPONSE

Accepted - payload is valid, scan in progress.

Response schema type: **object**

String status

BigID's Unmatched Data Coverage & Discoverability

Unmatched Data Coverage + Discovery Depth + Insights + Data Actions

100+ Agentless Native BigID Connectors Across Data Center & Cloud, Data-at-Rest & Data-in-Motion

MS-SQL · My-SQL · Oracle-SQL · Postgress · Informix · Sybase · Hana · Teradata · Vertica · Snowflake · Netezza · Hive · Redshift · Dynamo · S3 · EMR · Hadoop Hive · Aurora · Athena · Azure Blob · Azure File · GCP Blob · GCP Spanner · Elastic · O365 · GCP Big Query · Google Drive · One Drive · Slack · Gmail · Outlook · Exchange · Confluent / Kafka · DataBricks / Apache · API · Cassandra ServiceNow · CouchBase · MongoDB · Salesforce CRM · SAP ERP · Oracle ERP · SAP Datahub · SAP HANA · DB2 · VAX · Slack · SMB · NFS · CIFS · MS AD · Elastic · Sharepoint · Sharepoint Cloud · NetSuite · Workday · CosmosDB · Redis · Cassandra · Elasticsearch · Couchbase · BOX · Amazon S3 · Redhat OpenShift · IBM CP4D · TrustArc · VSAM

Custom BigID Connector

~6 weeks

New Connector Self-Service SDK

Now available

Generic API SaaS Connector

Now available

Any-SaaS Quick Connect

~Early Q2

Middleware Support

MuleSoft · DataHub · Denodo · Presto

Productivity & Messaging

- Office Docs
- Google Docs
- PDF
- Mail
- Chat

Binary & Bulk

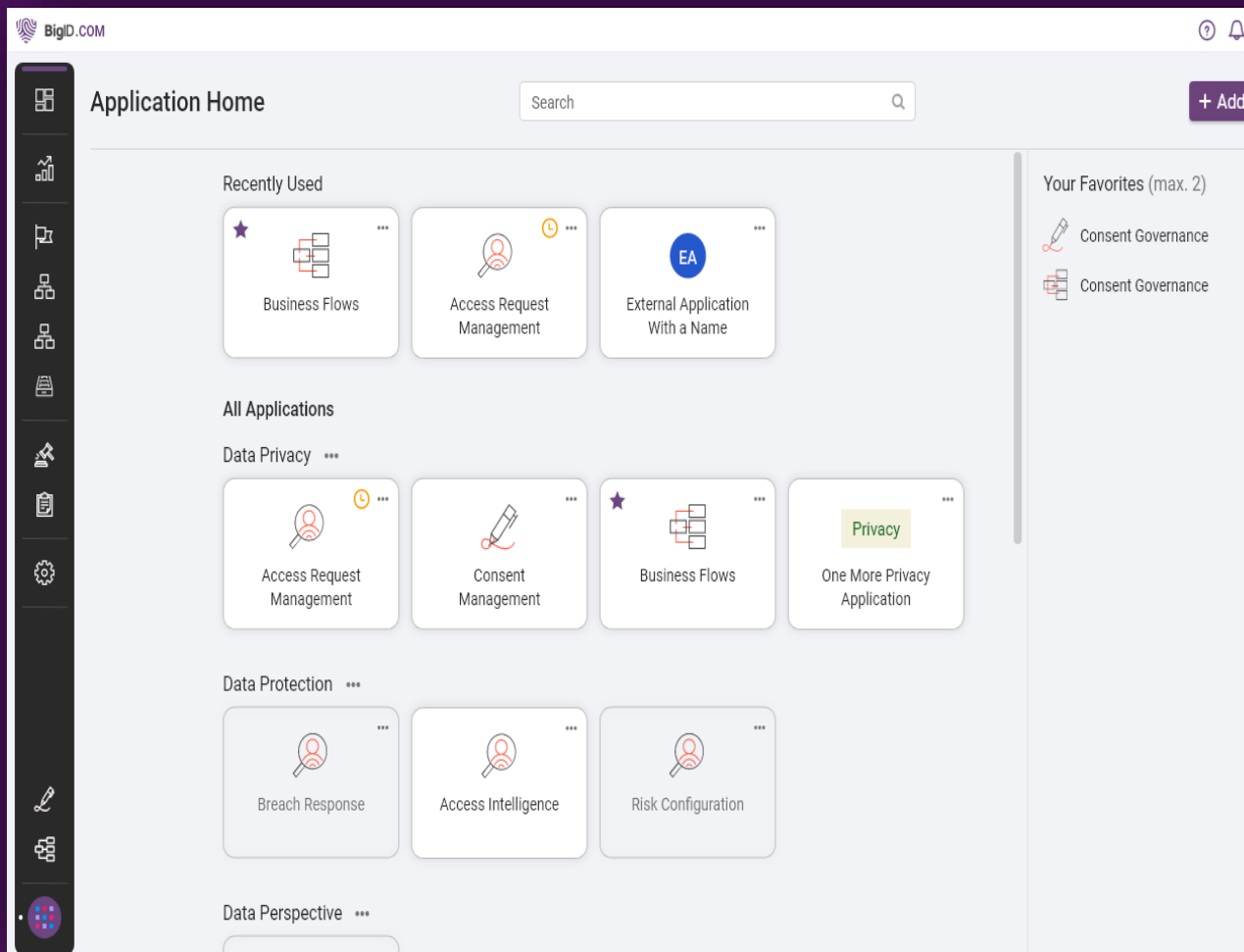
- Parquet
- Images / OCR
- Orc
- Avro
- Zip

Pipelines & Transports

- Kafka
- Kinesis
- API / Gateway
- Network
- FTP (Q2)

BigID App Framework

- **Build apps** for privacy, protection, and perspective
- **Extend BigID** discovery-in-depth to take action and solve problems
- **Get more value** from your existing tools



How to Use the BigID App Framework

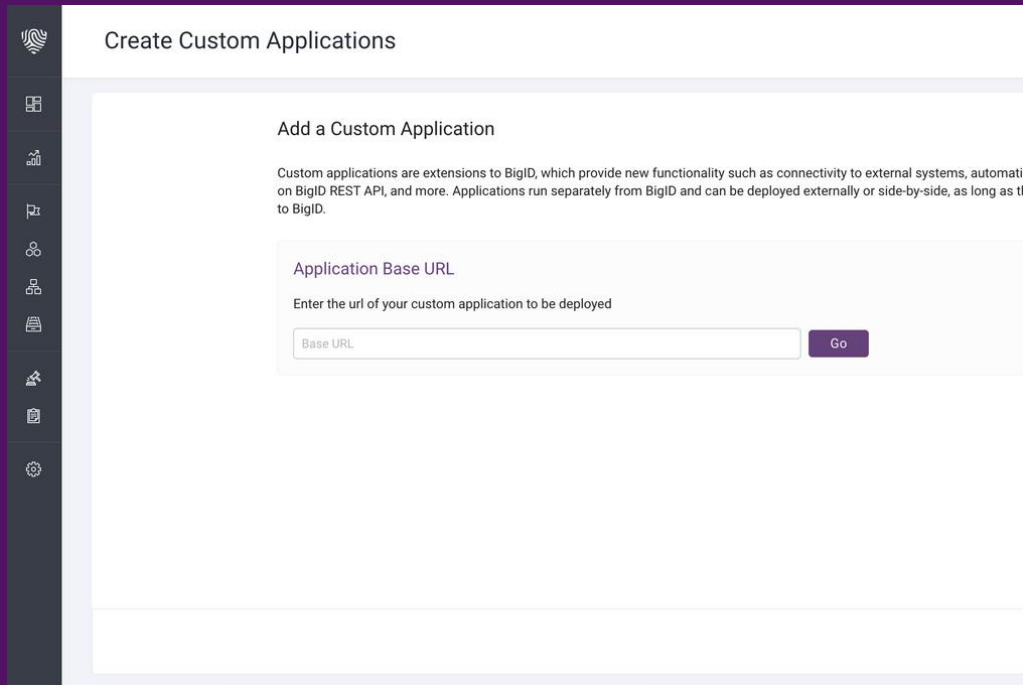
Make your data work for you

Quickly build your own Apps

Present custom perspectives
of your data

Automate data processes

Enhance your existing tools



The screenshot shows the 'Create Custom Applications' page in the BigID interface. On the left is a dark sidebar with a fingerprint icon at the top and a vertical menu of icons below it. The main content area has a white header with the title 'Create Custom Applications'. Below this is a section titled 'Add a Custom Application' with a descriptive paragraph: 'Custom applications are extensions to BigID, which provide new functionality such as connectivity to external systems, automation on BigID REST API, and more. Applications run separately from BigID and can be deployed externally or side-by-side, as long as they connect to BigID.' Below the text is a form with the label 'Application Base URL' and the instruction 'Enter the url of your custom application to be deployed'. The form contains a text input field with the placeholder 'Base URL' and a purple 'Go' button to its right.

Create Custom Applications

Add a Custom Application

Custom applications are extensions to BigID, which provide new functionality such as connectivity to external systems, automation on BigID REST API, and more. Applications run separately from BigID and can be deployed externally or side-by-side, as long as they connect to BigID.

Application Base URL

Enter the url of your custom application to be deployed

Go

But What About Data in Motion?

Data sharing - whether for services like back up or customer profile augmentation , cloud analytics - is increasingly facilitated through data streaming technologies (Kafka, AWS Kinesis)

Why is this a problem?

- **Lack of visibility** into data streaming technologies (Apache Kafka, AWS Kinesis) means blind spots for governing compliance (purpose of processing eg) for dev & analytics use cases
- Legal contracts exist only on paper, but **can't be validated** without insight into data movement

Why should customers care?

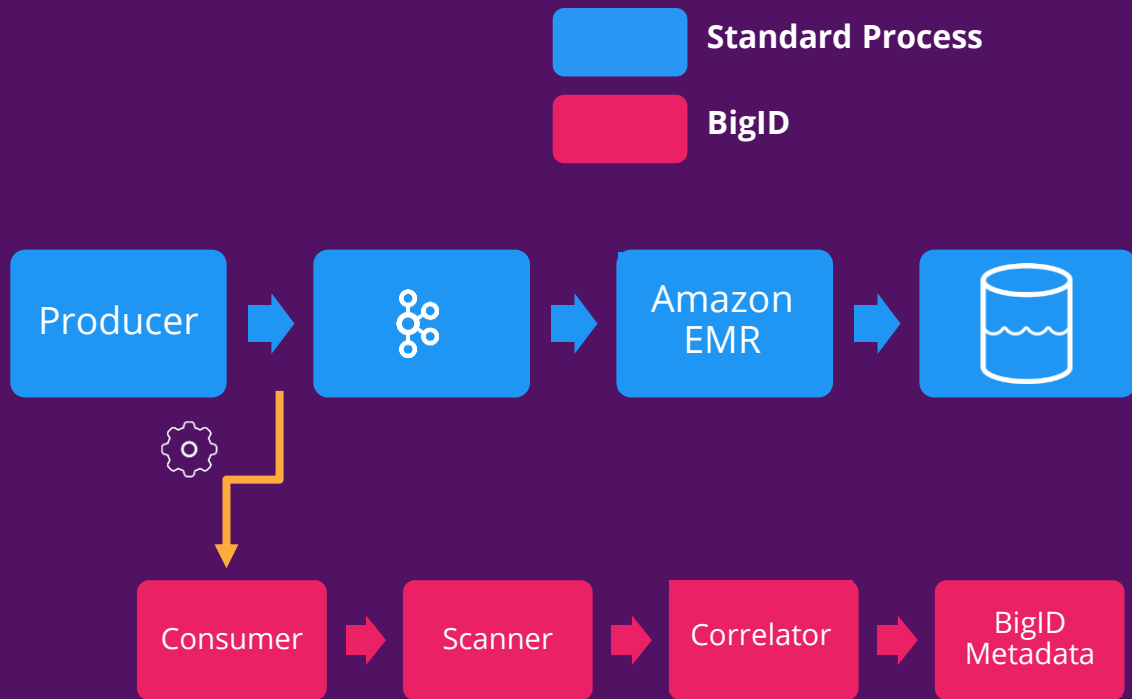
- Movement from transactional to continuous
- ~45% of data streaming workloads for basic data movement & processing ...
- Cloud streaming processing is growing, creating more blind spots across the data pipeline
- Data as a business asset = third party data is a key factor



USE CASE

Data in Motion Governance for Data Lakes

- Continuous Discovery of PI by payload
- Monitor the transmission of sensitive PII and PI via data streaming
- Policies for governance of consumer consent for data moving through the pipeline
- Alert on new PI found for remediation



Opt-Out & Do Not Sell

"A consumer shall have the right, at any time, to direct a business that sells personal information about the consumer to third parties not to sell the consumer's personal information.

This right may be referred to as the right to opt-out."

(Section 1798.120(a))

- **Important notes**

- Sale is defined broadly - can be an exchange of value, doesn't necessarily have to involve a monetary exchange
- Transfers between business affiliates can also be considered sales
- Must wait 12 months before seeking re-authorization to sell their personal information.

- **First Steps**

- Perform Data Mapping to determine how data is flowing
- Identify situations that constitute data sales
- Potential integration with vendor registries to evaluate what data is covered
- Integration with preferences management



DEMO



Key Contact Info

Alan Taylor
Global Alliance VP
alant@bigid.com

+1 214 930 6851 Cell
+1 214 548 5671 Home Office
Dallas Based

Tim Bayne
Global Alliances
Snr Sales Engineer
timb@bigid.com

+1 303 912 7011 Cell
Colorado Based

Sachin Khungar
Global Alliances
Technical Director
sachink@bigid.com

+44 7710 392809 Cell
London Based