EDM Webinar

From Data Discovery to Data Insight:
How to Leverage Automation and Machine Learning

Live Date: July 16, 2020

Featuring:
Nimrod Vax, Co-founder and Chief Product Officer, BigID
Peggy Tsai, VP of Data Solutions, BigID

Recording: bit.ly/2CgJ2dO
Presentation: bit.ly/330Sltu

edmcouncil.org   **bigid.com**

WEBINAR Q&A:

Can BigID provide the complete lineage of data management process / business process spread over multiple systems? (Example: Net worth of a customer from a BI report might be receiving data from multiple systems with complex calculations. Those systems could be geographically dispersed and some of data elements could be PIIs)

The ability to identify PII across systems that are geographically dispersed is an important part to being compliant to privacy regulations like GDPR. BigID currently assists in identifying the physical location and the stream of motion of data elements including BI reports and those systems with geographically dispersed data. BigID cannot currently identify the direction of the lineage flow yet, but integrates with lineage vendors in the market such as ASG and Manta to enrich their tools with classification, business technical and operational metadata to provide more holistic lineage information.

What hosting options does BigID provide?

BigID is an on-prem solution but can be deployed in any cloud such as AWS, Azure, or GCP. It can be deployed on Kubernetes and support auto scale and monitoring features. BigID can be delivered as a fully managed hosted solution by our partners.

Can it integrate SaaS solutions like Salesforce?

Yes. BigID has many out of the box integration solutions including Salesforce, Workday, NetSuite, SAP, Jira, and ServiceNow to name a few. Please see https://bigid.com/data-coverage/ for a full list of supported connectors.

Can BigID store technical data such as connection string to a source system and governed through a RBAC system (i.e. only authorized users can connect)?

Yes. BigID has a full RBAC model to restrict access to information stored on the BigID server. BigID also integrates with AD and SAML for authentication and SSO.

How does the licensing model work?

Please contact the BigID team directly at https://bigid.com/contact/

Who are the other players you compete for?

BigID's data intelligence platform is design for data privacy, security and data governance markets, along with an extensive partnership and integration network.

Does BigID support data discovery of email?

Yes. BigID supports data discovery in email including Microsoft Outlook Online, Microsoft Exchange, Google Email, and the full Google suite.

Can you do data discovery for unstructured data such as images, voice, web content, or text?

Yes. BigID's data discovery capabilities extend to all types of data, everywhere. Our unstructured data supports images, content, and text across Windows file shares, Unix/Linux file shares, SharePoint, Box, O365, Gdrive, Azure storage, AWS S3, GCP storage, and more. For a full list of support, please visit https://bigid.com/data-coverage/

Can you scan Teradata BTEK Scripts for data lineage using BigID?

BigID integrates seamlessly with specific lineage tools to provide a holistic lineage solution that starts with comprehensive discovery-in-depth.

Could you use this for a data discovery or eDiscovery search?

Yes. BigID's discovery-in-depth foundation automatically finds, identifies, and classifies all sensitive, personal and regulated data.

Since you monitor caches and accesses, could that be used reliably to determine who might have accessed a particular data source/directory/file change?

BigID currently provides a data access intelligence capability that collects metadata related to data source access, such as date last accessed, and can interrogate unstructured data such as file shares in the data center or cloud, that identifies open access permissions.

Do you consider AI/ML as a game changer technology for Enterprise Data Management and Governance? Or is it just another tool?

Yes. AI/ML is a game changer technology for Enterprise Data Management and Governance: by leveraging AI/ML techniques, organizations are able to get deeper insight, automate manual processes, apply more layers of understanding to their data, and get context from their data - improving data quality, time to value, and data intelligence.

How much training is needed to leverage AI/ML facilities within solutions? How big of a role does supervise learning play as compared to unsupervised learning?

BigID's solution employs both supervised and unsupervised learning to provide a holistic view of your data.

Can you handle data from popular applications like Office365 teams, SharePoint, OneDrive? How do you differentiate your product from eDiscovery software?

BigID supports and discovers data from Office365 - including OneDrive, SharePoint, and more (for a full list of support, see https://bigid.com/data-coverage/). BigID's data intelligence platform - and next-gen data discovery capabilities - is able to automatically discover, identify, relate, and classify more data across more data sources than traditional eDiscovery solutions.

Does this also provide the 360-degree view of data which it finds in motion?

Yes. BigID provides a number of data in motion connectors including Kafka and AWS Kinesis, along with a scan api you can use directly from applications, API gateways, and ETL tools. BigID gives a unified view of data in motion alongside data at rest.

Why is DCAM, DM in general not focusing that much on production process data? The amount of collected data (from edge, mobile devices, machines, cars) is huge.

DCAM, as supported by the EDM Council, is designed as a best practice framework across the entire enterprise data supply chain to include all types of data both structured and unstructured. Product process data which includes data collected in different formats is often unstructured and non-standardized. Therefore, it may be difficult to relate back to the other data sources in the organization. BigID supports the management and governance of all these data sources that should be held in high regard - so that you can improve data quality, data management, and data insight for all types of data.

How do you find the data? I suppose you need a direct access to the Data Source?

BigID is agentless and uses native protocols to connect to the data sources: it does not copy or move the data.

What do you use for continuous data pipelining?

BigID integrates with Kafka and Kinesis to scan, manage, and monitor data pipelines.

What is the role (and requirement) for human review and resources for reviewing recommendations (confidence intervals, false positives, etc.)?

The BigID data intelligence doesn't require human review or manual alignment, instead leveraging NLP and ML techniques to automatically and accurately discover, classify, and map data across both unstructured and structured sources. Our ML-techniques result in fewer false positives and higher accuracy, by layering multiple modes of analysis, discovery, and insight. Organizations have the option to manually review the matches, adjust sensitivity scores, and retrain the model as needed.