



EDM Webinar

Data Governance for Generative AI

A conversation with...



Peter Wang,
Staff Product Manager,
Databricks



Jocelyn Houle
Sr. Director, Product
Management,
Securiti



Moderator



Eric Bigelsen
Head of Industry
Engagement
EDM Council



Jocelyn Houle
Sr. Director, Product
Management
Securiti



Peter Wang
Staff Product Manager
Databricks



One New Variable of GenAI Impacts Almost Everything!

GenAI's Probabilistic Foundations

of Upcoming AI Agents for Knowledge Creation, Automation & Autopilots

Use of Data & Resources

Generation of Data &
Knowledge

Interaction with
Customers & Service
Delivery

It affects many internal Data Controls.

New Level of Visibility and Controls at Every Layer of Data Stack

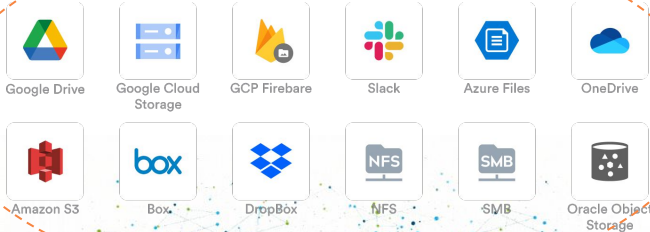
Data + AI will drive major transformations in human history

Your Proprietary Data Anywhere

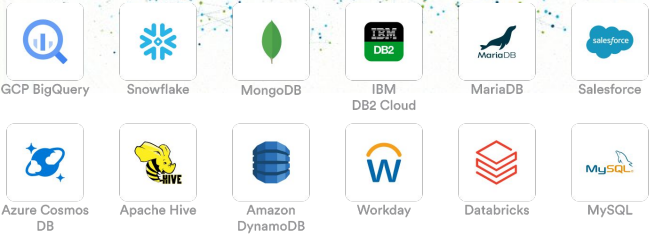
GenAI in Databricks

GenAI Apps & Innovation

Unstructured Data

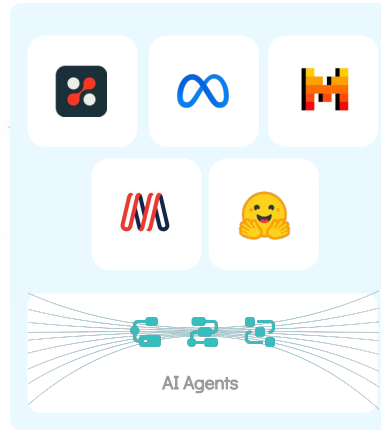


Data-Command Graph Single Source of Truth



Structured Data

AI Models + Agents

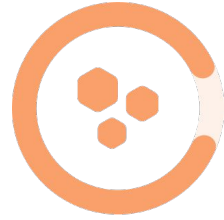


Copilots & Assistants

Business Process Automations

New Category of Apps

Insights & Analytics



90% of organizational data is now

unstructured, and must

be safely governed when used by GenAI models.

Top 10 Considerations for Enabling Safe Use of Unstructured Data with GenAI



1. Discovery of Unstructured Data

Discover unstructured data assets across diverse storages.



2. Cataloging of Unstructured Data

Catalog all files and objects that can be used for genAI projects.



3. Classification of Unstructured Data

Classify all files and documents based on their sensitivity and other attributes.



4. Entitlements of Unstructured Data

Preserve enterprise entitlements at source systems for genAI projects.



5. Lineage of Unstructured Data

Assess and document the origins and uses of data in genAI projects.



6. Curation of Unstructured Data

Curate and auto-label files and objects for use in genAI projects.



7. Extraction of Unstructured Data

Extract data from hundreds of unstructured formats to enhance data utilization.



8. Sanitization of Unstructured Data

Sanitize / redact / mask / sensitive data from files for use in genAI projects.



9. Quality of Unstructured Data

Ensure files are fresh, unique, and relevant before feeding to GenAI models.



10. Security of Unstructured Data

Secure unstructured prompts, retrievals, and responses with distributed, context-aware LLM firewalls

Poll

Is unstructured data like documents, videos and sound files part of your data estate?

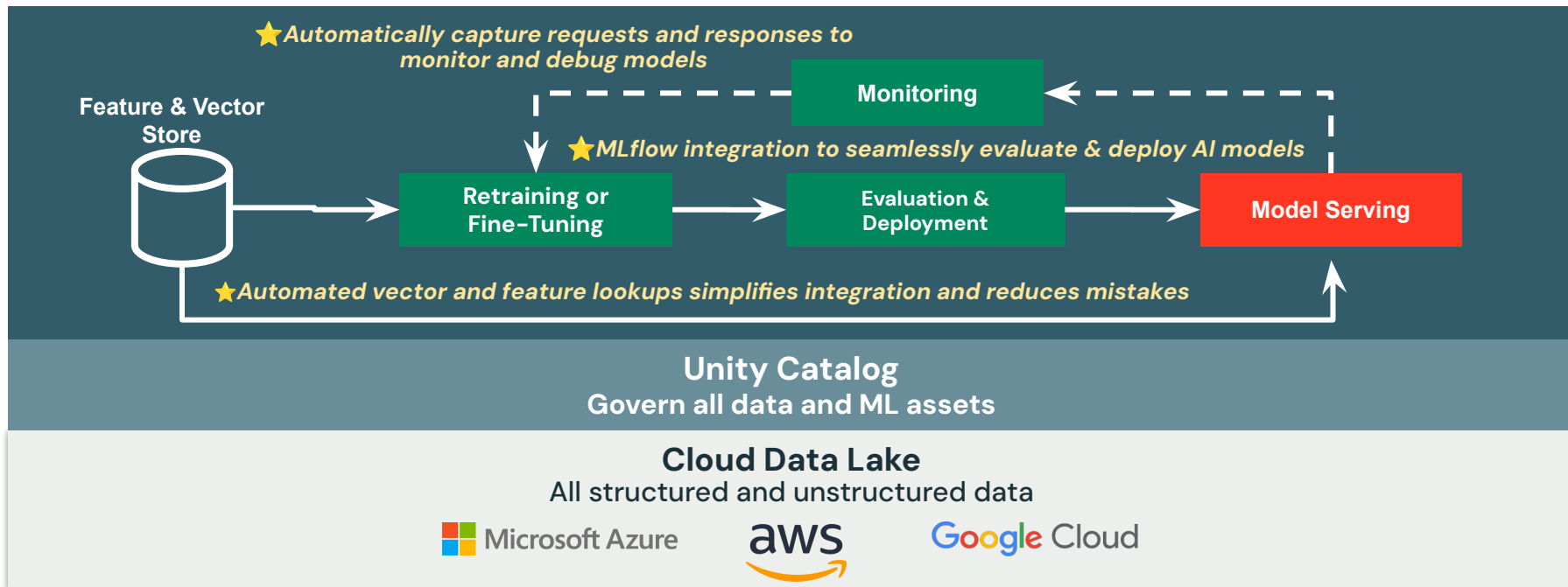
- Yes - it is a big part
- No - we are pretty standard on structured
- I don't know

Build AI Apps with Model Serving

Low TCO

Accelerated deployments

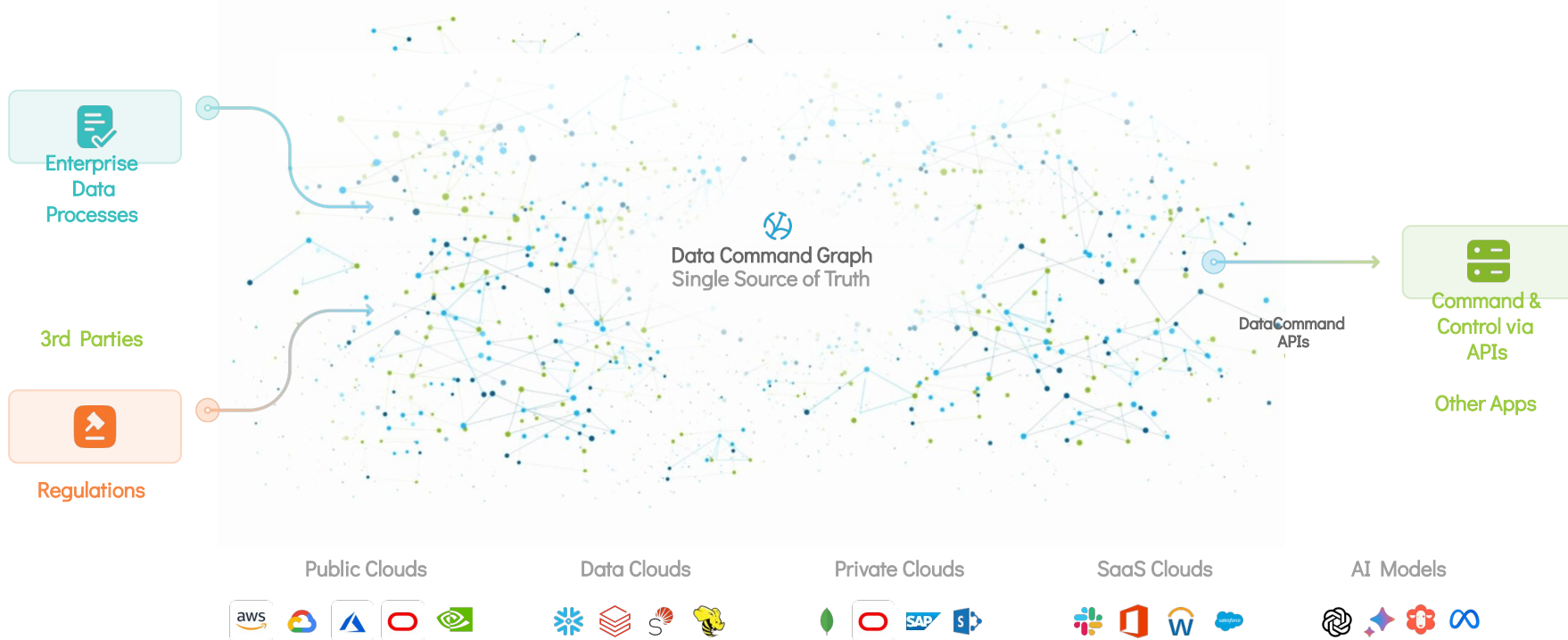
Fewer Mistakes



Data Command Graph provides deep contextual insights for Unstructured + Structured Data

A baseline need for GenAI pipelines & use cases

Data



Data Command Graph provides deep contextual insights for Unstructured + Structured Data

A baseline need for GenAI pipelines & use cases

Data

- **Data Command Graph** is a Knowledge Graph that automatically captures all key metadata and relationships between them:
 - Data Systems
 - Files/ Folders
 - Tables/ Columns
 - Data Sensitivity
 - Access & Entitlements
 - Internal Policies & Controls
 - Applicable Regulations
 - Related AI Models / Pipelines
- Advanced out-of-the-box Data Classifiers to auto-classify sensitive data. Advanced NLP and EDM techniques for unstructured and structured data.
- Answer any question around Data+AI and their relationships with a multidimensional graph model. This enables the safe use of data to go from POCs to Production.

Find answers to your key questions

What sensitive data exists?



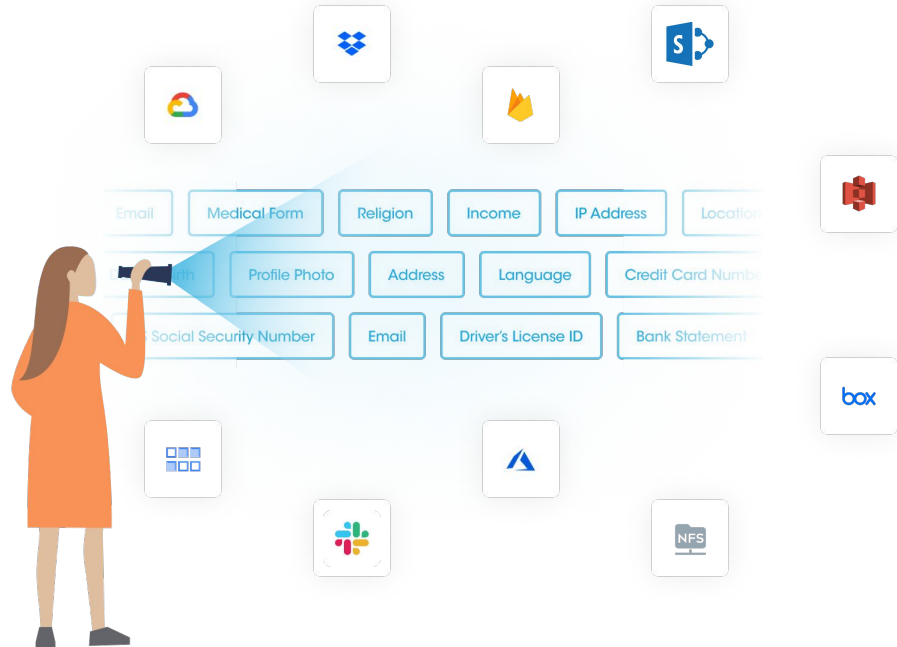
Unstructured Data takes the center stage with GenAI!

Contextual Insights for Unstructured Data from all key perspectives, a baseline need for Safe GenAI Pipelines & Use Cases

- Files of all types (docs, audio, video, images etc.). CLOBs.
- File categories (legal, finance, HR etc.) based on content
- Access and user entitlements
- Sensitive objects within a file
- Regulations applicable on file content
- File quality (freshness, relevance etc.)
- **Lineage of files & embeddings** used in GenAI pipes

Curate data by analyzing content and automatically adding **Data Labels** to files based on content

Empower the use of Unstructured Data for **LLM Training/Tuning, RAG** and all related GenAI Apps



Comprehensive Data Intelligence for Structured Data

Contextual Insights for **Structured Data** from all key perspectives, empowering AI Agents & AI Use Cases

- Tables, Columns, Schemas
- Business Context & Glossary
- Access and user entitlements
- Sensitive Data Columns & Elements
- Regulations applicable to Data
- Quality
- Lineage

Structured Data, combined with Unstructured Data enables variety of GenAI use cases, AI Agents and GenAI Apps.



Poll

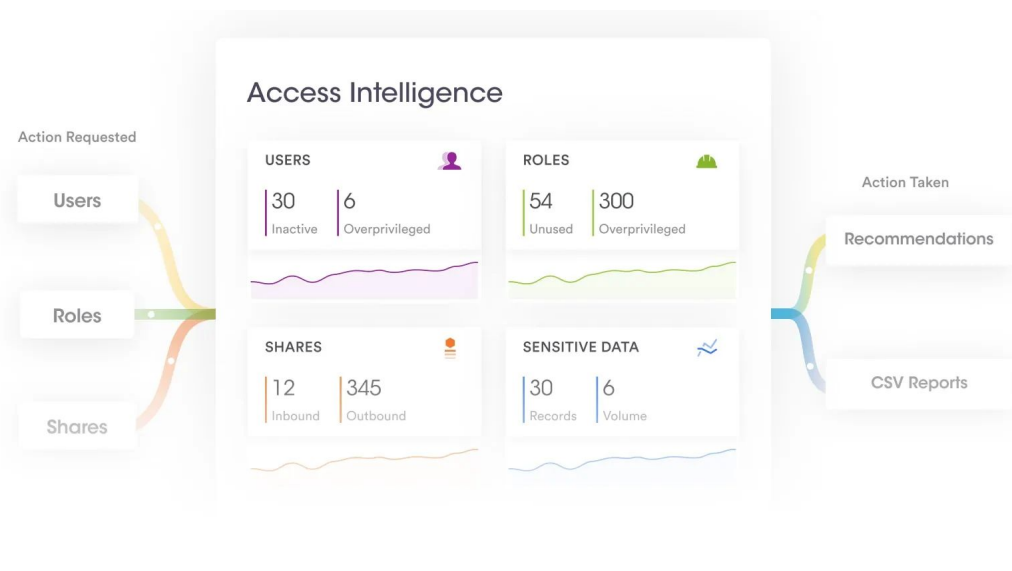
Do you need granular data access down to the row, column or element level?

- Yes
- No
- I don't know

Data Access & Entitlement Intelligence empowers safe GenAI pipelines

Honor data entitlements to prevent data leakage and ensure safe data utilization

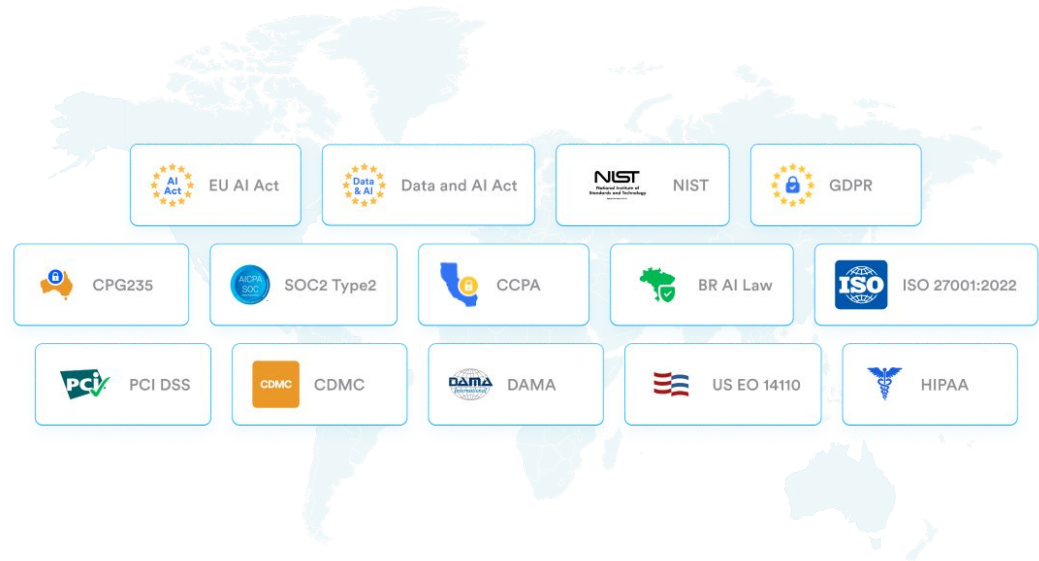
- Identify users and roles with access to
 - **Unstructured data** - buckets, folders, and files or documents in data lakes and file shares
 - **Structured data** - databases, tables, and sensitive data in data warehouses (like Databricks)
- Map the relationships between users, roles, and the permissions they have for accessing documents or datasets.
- Enables Dynamic Masking of Structured Data for Certain Roles and Users, for safe use with AI.
- Entitlement information can be used within GenAI pipelines to enforce same entitlements at the prompt level.



Regulatory context of data enables its compliant use

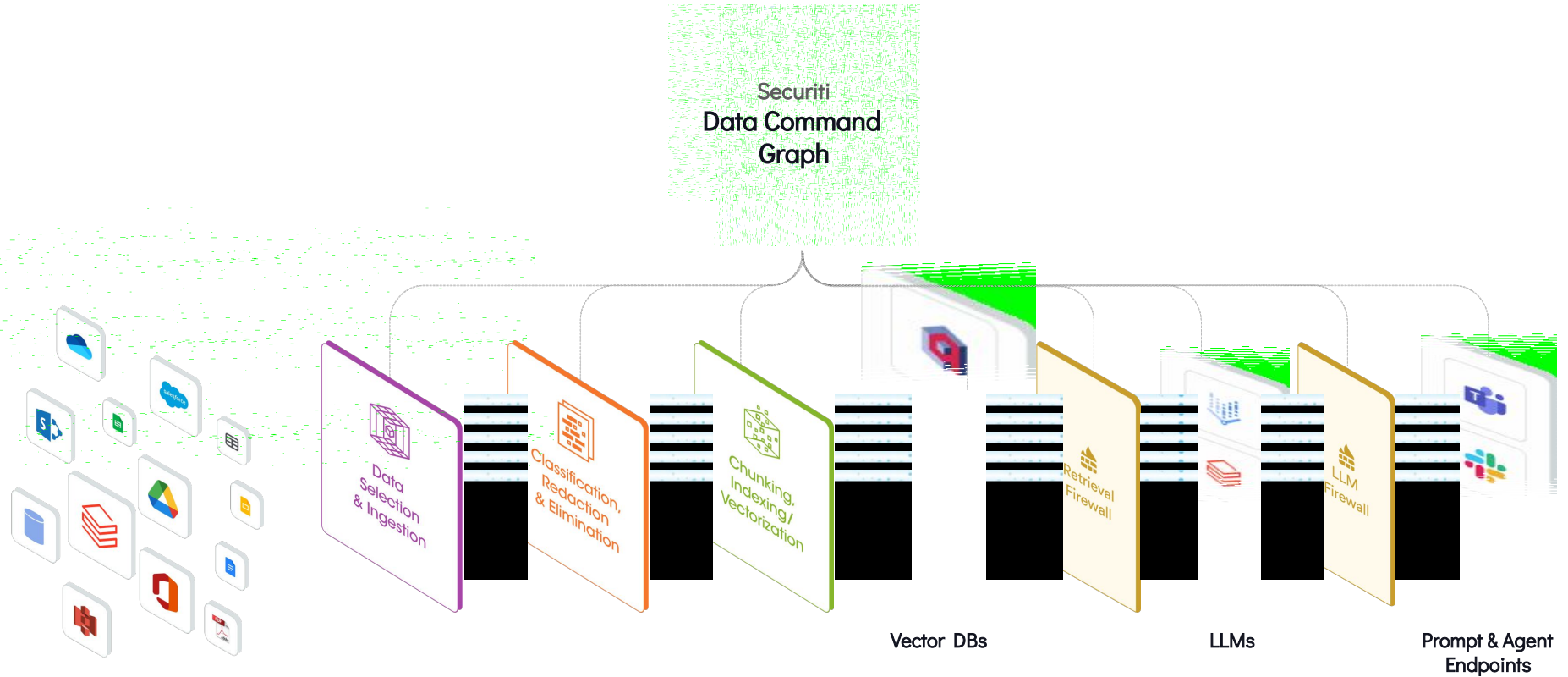
Gain insights into regulatory risks based on applicable Data+AI laws

- Categorize data based on sensitivity levels (as per GDPR, HIPAA, and other data protection regulations) and determine what data should be anonymized or redacted before feeding into AI models.
- Identify and mitigate risks related to data privacy, security, and other regulatory concerns.
- Enable the compliance of GenAI projects with applicable AI Regulations (EU AI Act, NIST AI RMF) and Data Regulations (GDPR, CPRA, GPD etc).



Safe Enterprise AI with Databricks & Securiti AI

Building Blocks of a GenAI Pipeline

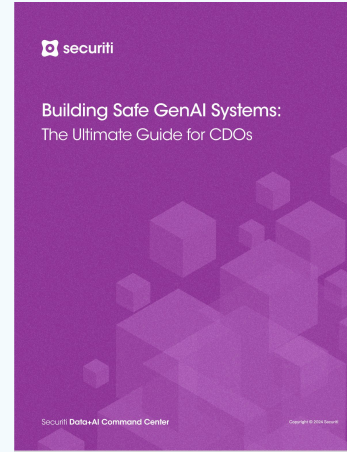
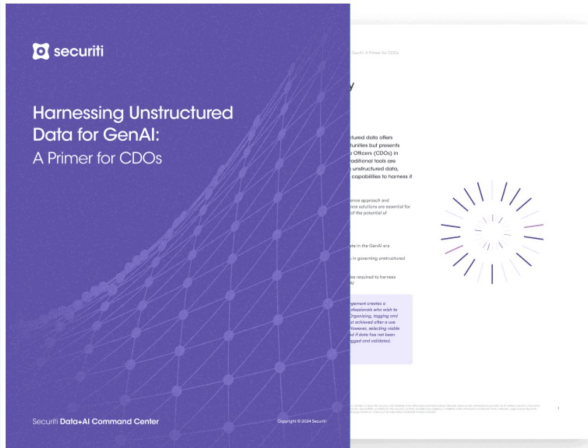


Questions?



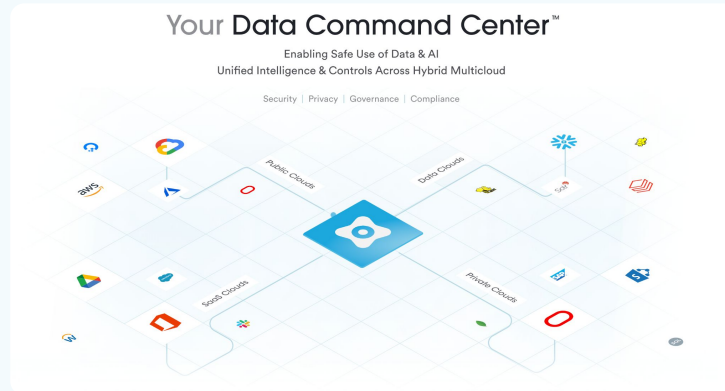
EDM Webinar 

For more information, please check out these resources.



Scan Now

Content Bundle



Join EDM Council and our membership community of companies...



The screenshot shows the EDM Council website homepage. At the top, there is a navigation bar with links for Membership, Frameworks, Training, Engage, Innovation, About, Sign in, and Join now. The main header features the EDM Council logo and a large blue banner with the text "Global Advocates for Data & Analytics Management". Below this, there is a sub-header: "The leading global trade association providing best practices, standards and education to data and business professionals in our data-driven world." A "What we do" button is visible. To the right, there is a "TODAY'S HIGHLIGHTS" section with three news items: "Bank of Valletta becomes the newest member to join the EDM Council", "EDM Council welcomes Webber Wentzel as its newest member", and "Lion Group joins EDM Council as its newest member". At the bottom of the screenshot, there is a section titled "Join a vibrant community of 25,000+ business leaders, CDOs, and data and analytics professionals across all industries." followed by logos for Bank of England, NOVARTIS, HSBC, AWS, Schneider Electric, Microsoft, Google, and AEGON. Two buttons are present: "Explore membership" and "See all 350+ member organizations".



350+ Member Firms

Cross-industry,
including Regulators



25,000+

Professionals



Worldwide

Americas, Europe,
Africa, Asia, Australia

edmcouncil.org



EDM Webinar 

Thank you!

**FOR MORE INFORMATION,
CONTACT:**

info@securiti.ai

OR VISIT ONLINE:

www.securiti.ai

