Data Discovery and Classification:

- Connected to AWS, Azure, M365 and Snowflake
- Discovered 842 datastores & scanned **586**
- Discovered **1.2T** records into **99** data classes
  - **Financial**: Credit Card Number , IBAN,Revenue, Card Verification Code
  - **PII:** US SSN, Taxpayer ID**,** Passport Number, Aadhaar Number, Drivers License
  - **Health:** Patient ID, Medical Condition
  - **Business & IP:** Company Name, Company Address, Employer ID Number
- 33% (33 out of 99) of the data classes in Ahead / Capgroup sandbox was classified **using ML models.**
  - Transaction ID , Payment ID, Credit Card Statement , Order ID, Bank Statement, Resume, W2, NDA, Client ID
- **Context**
  - Invalid Credit Card Numbers
  - Invalid IBAN Numbers
  - Synthetic Data
- **Data Subject Residency**
  - India
  - California
  - Canada

In Snowflake we Classified using without column names

Actionable Insights into risk

- Credit Card Numbers in Plain Text
- 34 ghost / stale data
- Sensitive data is unencrypted in Oracle
- EC2 hosting a Database containing sensitive data
- 114 stale identities with access to sensitive records
- 329 Power Users with some with access to restricted and confidential data

M365 Specific findings:

1. **Identified sensitive records in ghost drives**
2. 117 Files Shared Publicly
3. 216 files shared across the organization
4. Files shared with personal email
5. M365 Ghost Drives contain sensitive information