AI-POWERED EDR: STREAMLINING BLACKBERRY CYBERSECURITY WITH DATABRICKS

Justin Lai – Distinguish Data Architect
Robert Lombardi – Director of Product Management
Digan Parikh – Sr. Solution Architect (Databricks)

BlackBerry®

Challenges in Cybersecurity

4 common challenges in Cybersecurity operations

Cost

Scale

ML & AI

Centralization

Databricks Vision for Cybersecurity

Databricks for Long-Term Storage + Analytics integrating with other tools like SIEMs & SOARs

Operations

Engineering

Risk & Compliance

SIEM | **SOAR** | **Intel** | **Fraud, AML, Insider Threat** | **Fusion Analytics**

Databricks Data Intelligence Platform

Data Sources

OS	Security	Agent	Network	PaaS/SaaS
(Windows, Linux)	(XDR, Antivirus)	(Antivirus, DLP)	(IPS, Sandbox)	(CSP, SSO, AD)

ROBERT LOMBARDI

Director of Product Management

- Started as an Engineer @ BB in 2009
  - Platform & Product Design
  - Advanced Technology & Research
- Transitioned to Product Management in 2017
  - Consumer Security Applications for Mobile
  - Launch of Protect Mobile (MTD)
  - Cylance Console Platform
  - CylanceOPTICS
  - CylanceENDPOINT

JUSTIN LAI

Distinguished Data Architect

- Started at BlackBerry 2010
- Worked in the security field for the past 12 years
- Data Platform Team for over 4 years
- Second time presenting at DAIS

BLACKBERRY CYLANCE

Cybersecurity Portfolio Overview

CylancePROTECT (EPP)

- Best-in-class Antimalware AI
- Attack Surface Reduction
- Exploit Prevention

CylanceOPTICS (EDR)

- MITRE ATT&CK
- Telemetry Capture & Context
- Incident Response

CylanceGATEWAY (ZTNA)

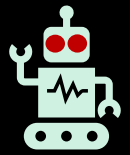
- Secure Network Communications
- Network Detection & Response
- Content Filtering & Firewalling

CylanceMDR

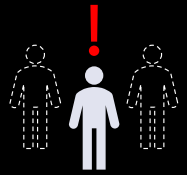
- Managed Detection & Response
- Managed Threat Hunting
- Environment Tuning
- Secure Critical Communications
- Incident Response
- Digital Forensics
- XDR Enabled

OPPORTUNITIES

Alert fatigue



Fear of AI



Not enough resources

Security complexity

Not the right resources

Evolving threat landscape

DESIRED OUTCOMES

Reduce the burden and need to perform monotonous high friction tasks so that customers can take remediation steps confidently and efficiently

Support & enable customers to leverage AI where needed and gradually grow confidence in the system with transparency

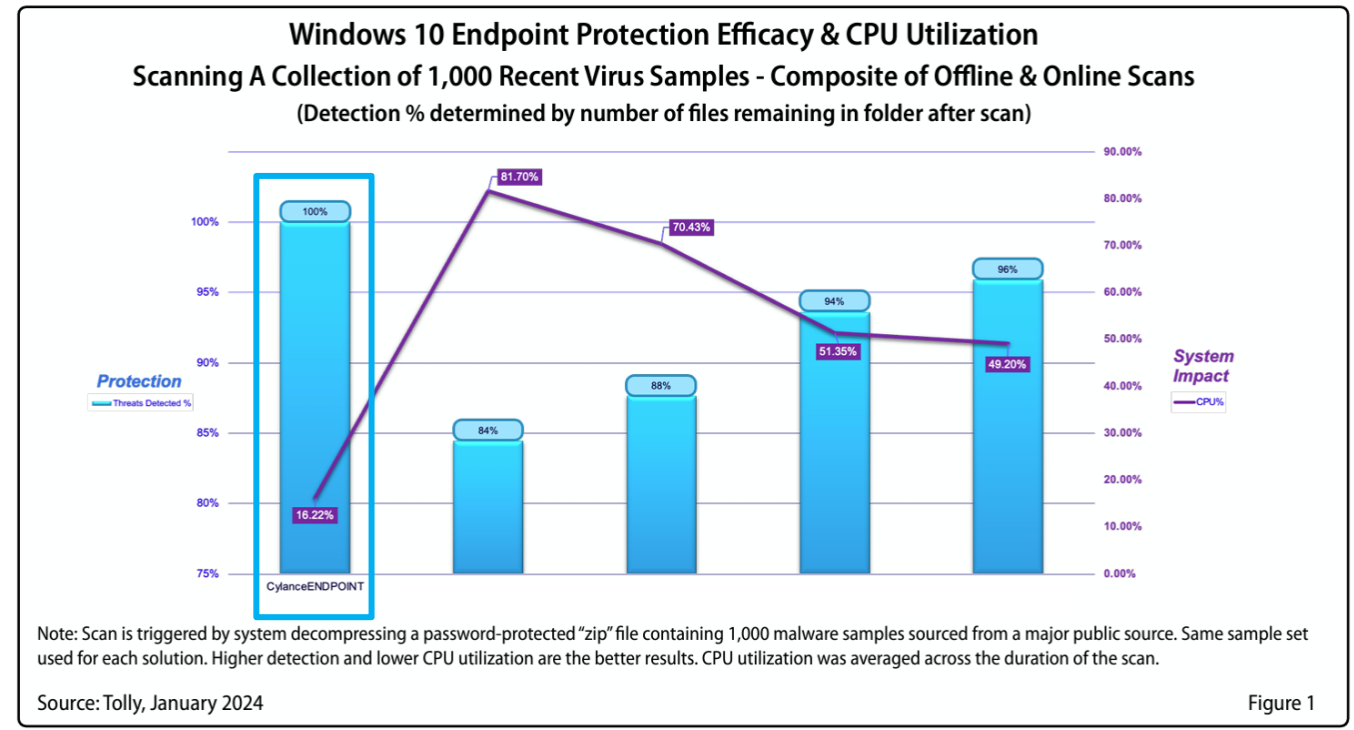
Enable upskilling security operations centers of all sizes by providing clear context & guidance throughout the response workflow

Provide focus and clarity to end users so that they can spend their very limited time on high value tasks

CYLANCE AI

Innovation in *Predictive* & Generative AI

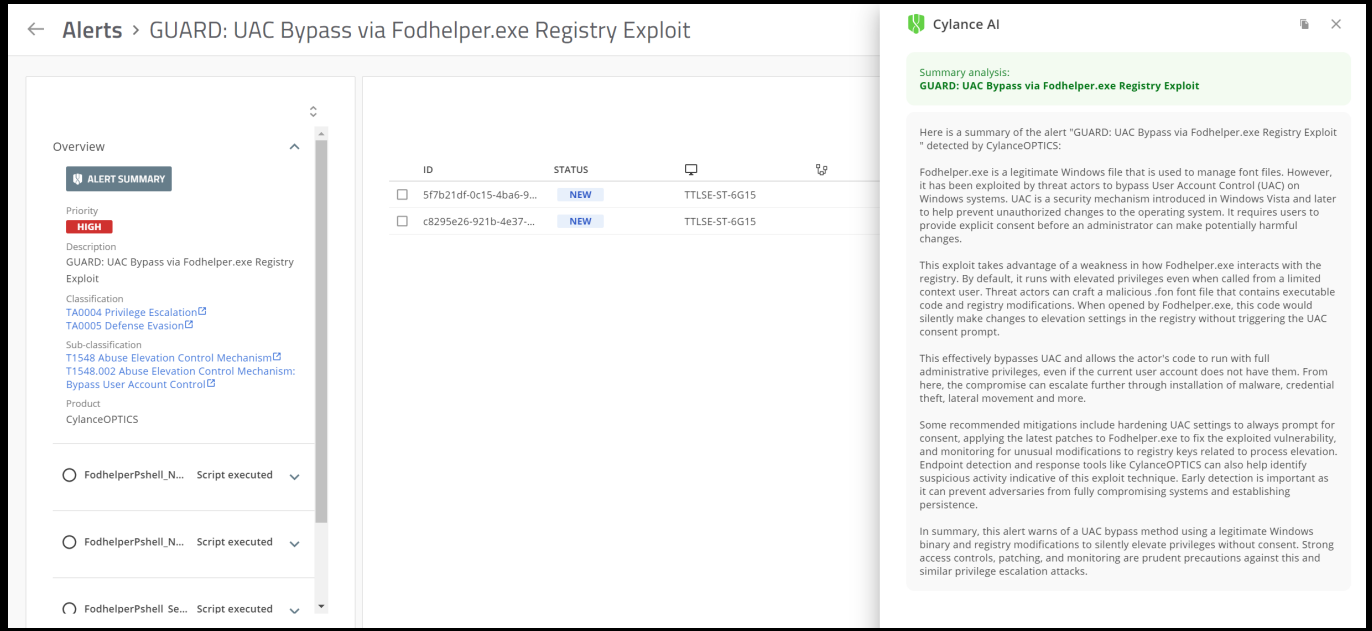
- **Threat Prevention**
- SOC Assistant
- Kill-chain Summarization
- AI-Assisted FP Handling
- Automated Tuning
- Adaptive Policy
- Incident Generation



CYLANCE AI

Innovation in Predictive & *Generative* AI

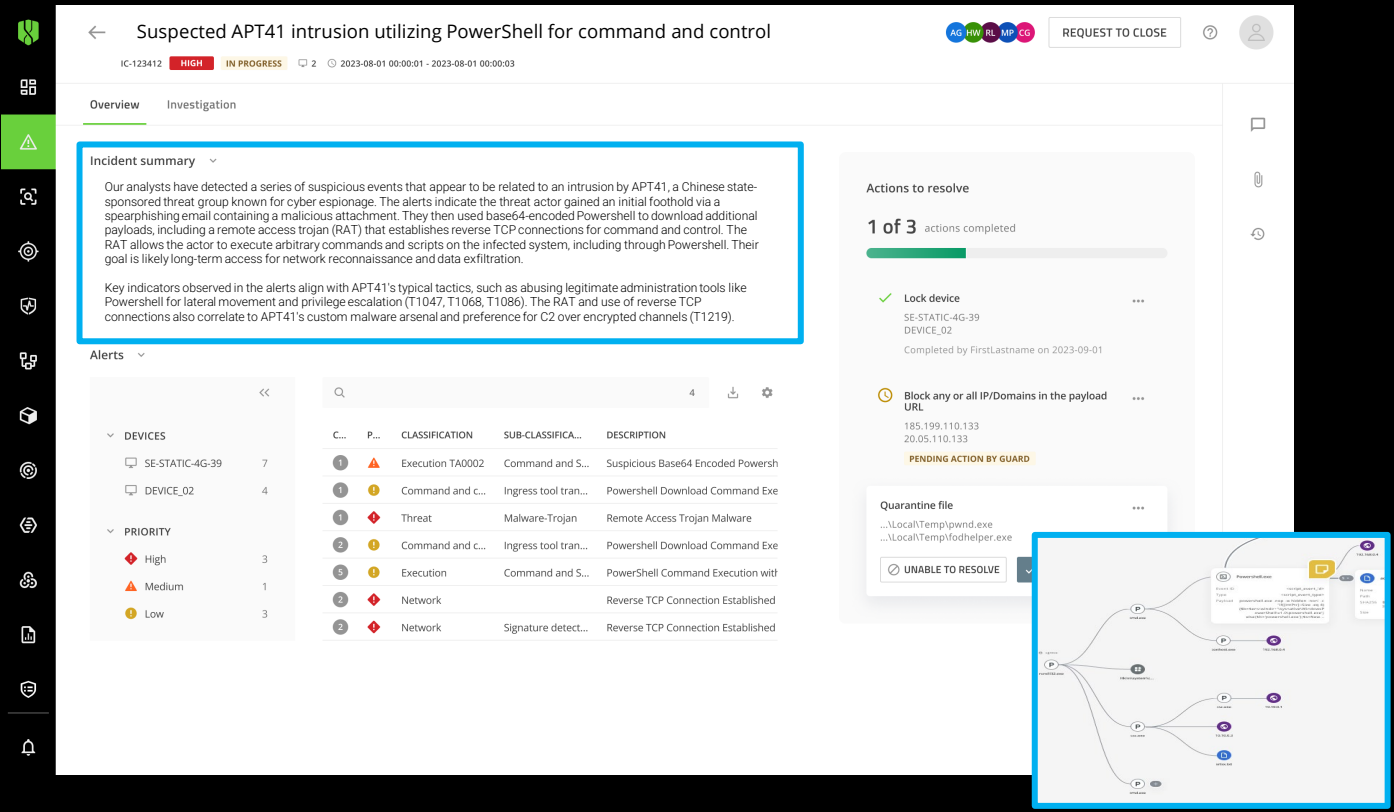
- Threat Prevention
- **SOC Assistant**
- Kill-chain Summarization
- AI-Assisted FP Handling
- Automated Tuning
- Adaptive Policy
- Incident Generation



CYLANCE AI

Innovation in Predictive & *Generative* AI

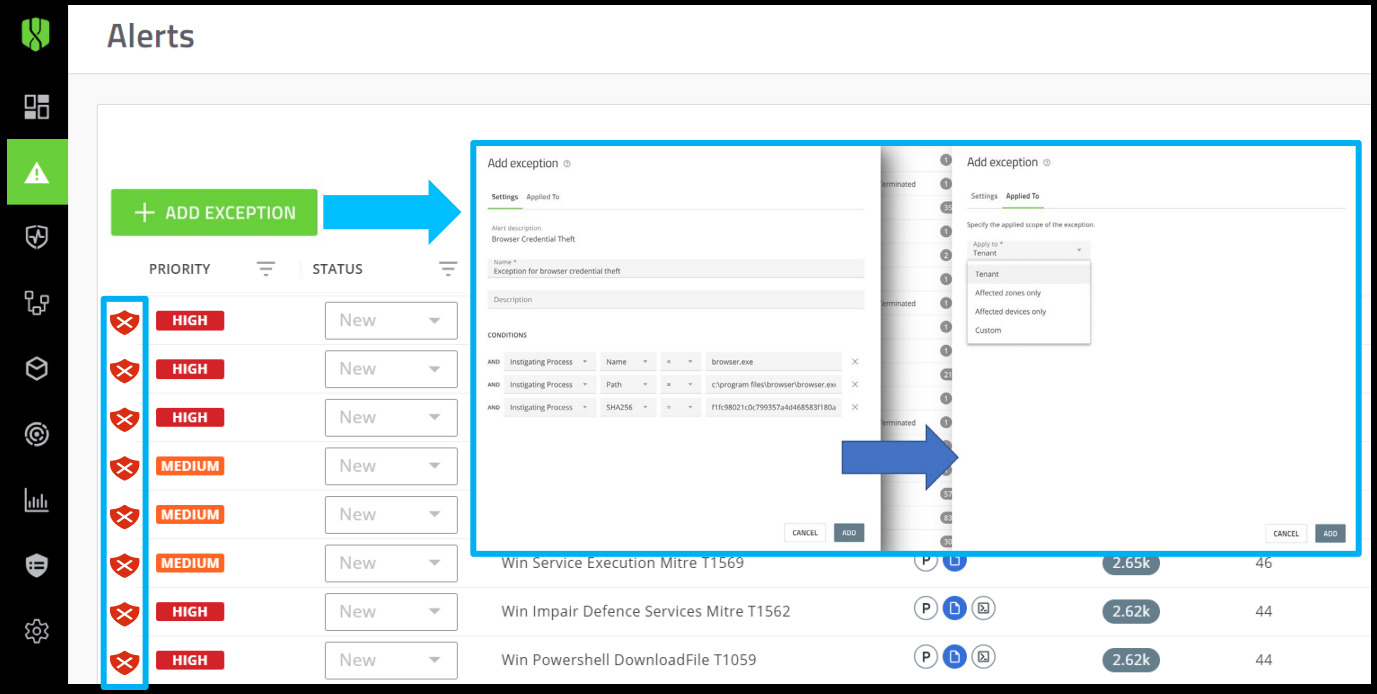
- Threat Prevention
- SOC Assistant
- **Kill-chain Summarization**
- AI-Assisted FP Handling
- Automated Tuning
- Adaptive Policy
- Incident Generation



CYLANCE AI

Innovation in *Predictive & Generative AI*

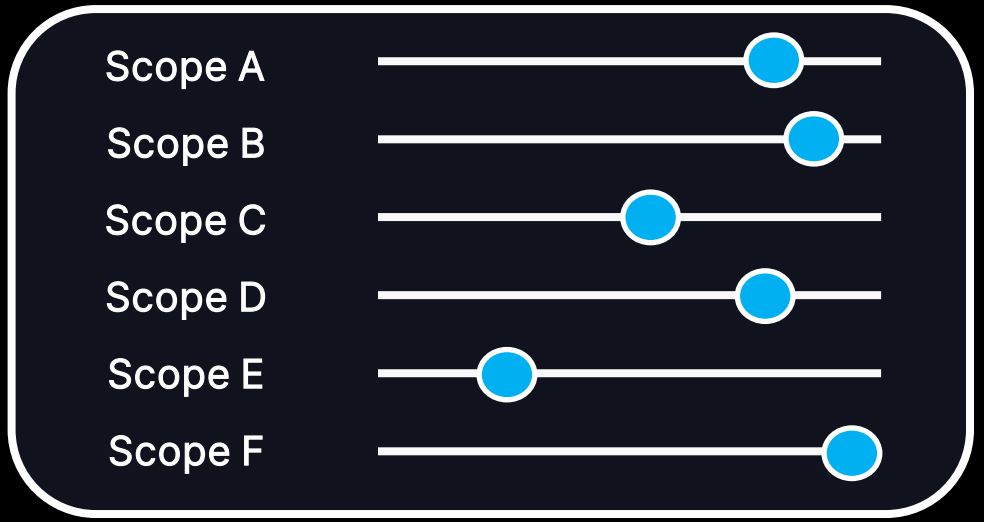
- Threat Prevention
- SOC Assistant
- Kill-chain Summarization
- **AI-Assisted FP Handling**
- **Automated Tuning**
- Adaptive Policy
- Incident Generation



CYLANCE AI

Innovation in *Predictive* & Generative AI

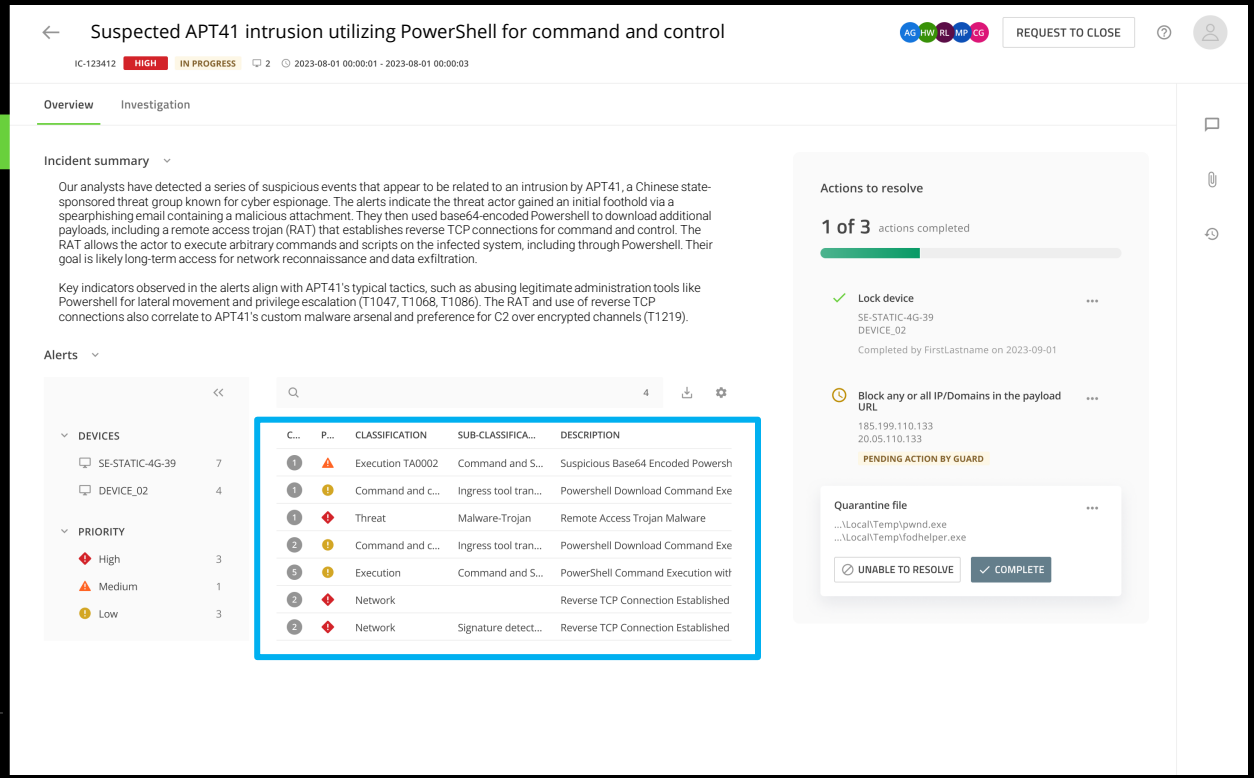
- Threat Prevention
- SOC Assistant
- Kill-chain Summarization
- AI-Assisted FP Handling
- Automated Tuning
- **Adaptive Policy**
- Incident Generation



CYLANCE AI

Innovation in *Predictive & Generative AI*

- Threat Prevention
- SOC Assistant
- Kill-chain Summarization
- AI-Assisted FP Handling
- Automated Tuning
- Adaptive Policy
- **Incident Generation**

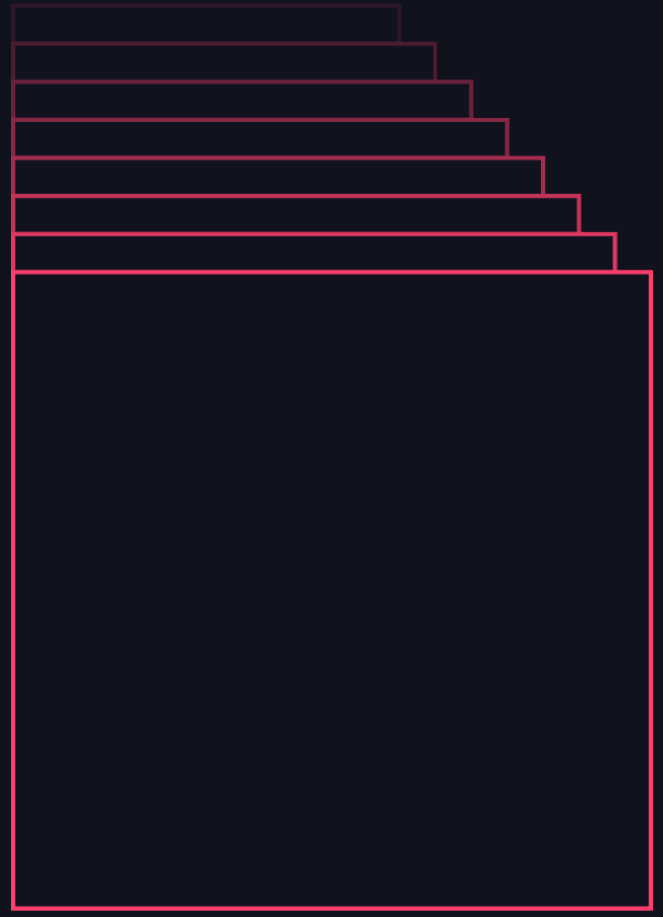


AGENDA

- High Level Overview
- Leveraging AI in our day to day
- Machine Learning Use Cases
- The Future

DATA+AI SUMMIT BY databricks

HIGH LEVEL OVERVIEW



Data Lake to Lake House to Intelligence Platform

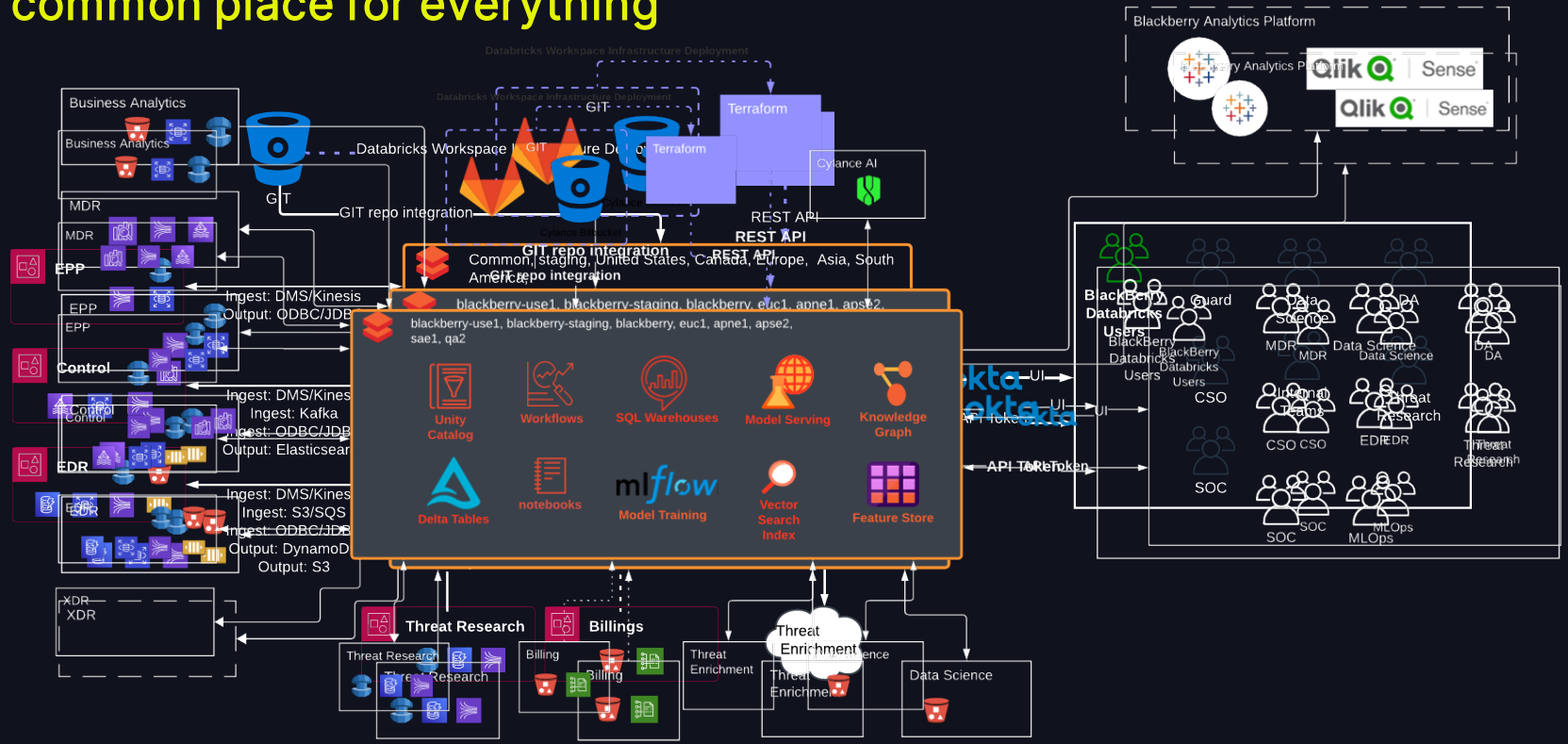
HIGH LEVEL OVERVIEW

Data lake to lakehouse to intelligence platform

- Partnered with Databricks to support our EDR Infrastructure
- Built a data lake to support our EDR product
- Built a cybersecurity data lakehouse
- Building a data intelligence platform

HOW DOES IT LOOK?

A common place for everything



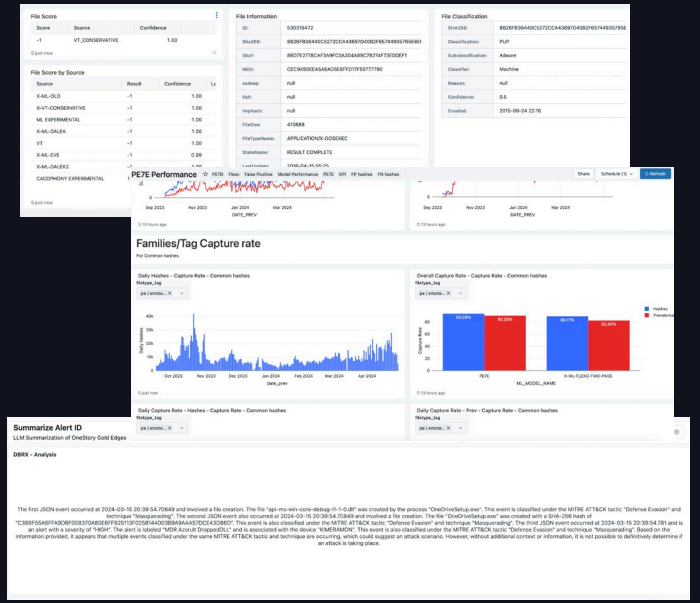
USE CASES

Data Lake to Lake House to Intelligence Platform

DASHBOARDING

USE CASES

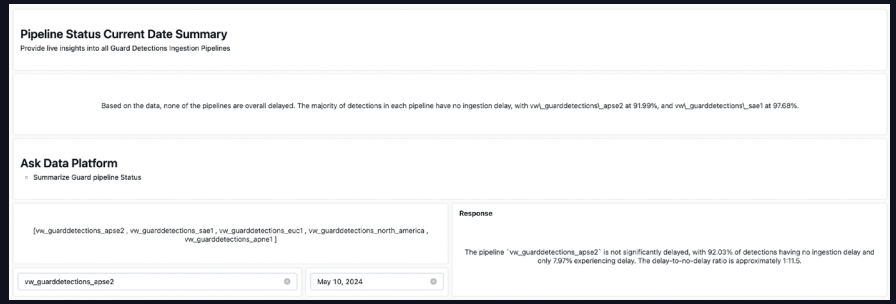
- Visualize how our models are performing globally
- Enables collaboration between engineers from different area of speciality
- Low code way to concept up visual elements powered by AI



AI IS PART OF OUR DAY TO DAY

Use Case: Inferring Pipeline Status

- Generate Table Aggregates
- Leverage SQL to prep data for LLM to process
- Visualize LLM output in an easy to use format
- Leverage Dashboard inputs to enable interaction with LLM




```
select
  concat(
    "For the pipeline ",
    a.view_name,
    ", there are ",
    a.detection,
    " detections with ",
    classification,
    " and represents ",
    string(((a.detection / sum(a.detection) OVER())) * 100),
    "% of total detections.",
    CASE
      WHEN classification = "no ingestion delay"
      AND ((a.detection / sum(a.detection) OVER())) * 100 >= 90 THEN "Pipeline is not suffering any delay."
      WHEN classification = "no ingestion delay"
      AND ((a.detection / sum(a.detection) OVER())) * 100 <= 90 THEN "Pipeline is delayed."
      WHEN classification = "ingestion delay"
      AND ((a.detection / sum(a.detection) OVER())) * 100 >= 100 - 90 THEN "Pipeline is delayed."
      WHEN classification = "ingestion delay"
      AND ((a.detection / sum(a.detection) OVER())) * 100 <= 100 -90 THEN "Pipeline is not delayed."
    END
  ) as summary
```

STREAMING EDR MODEL INFERENCE

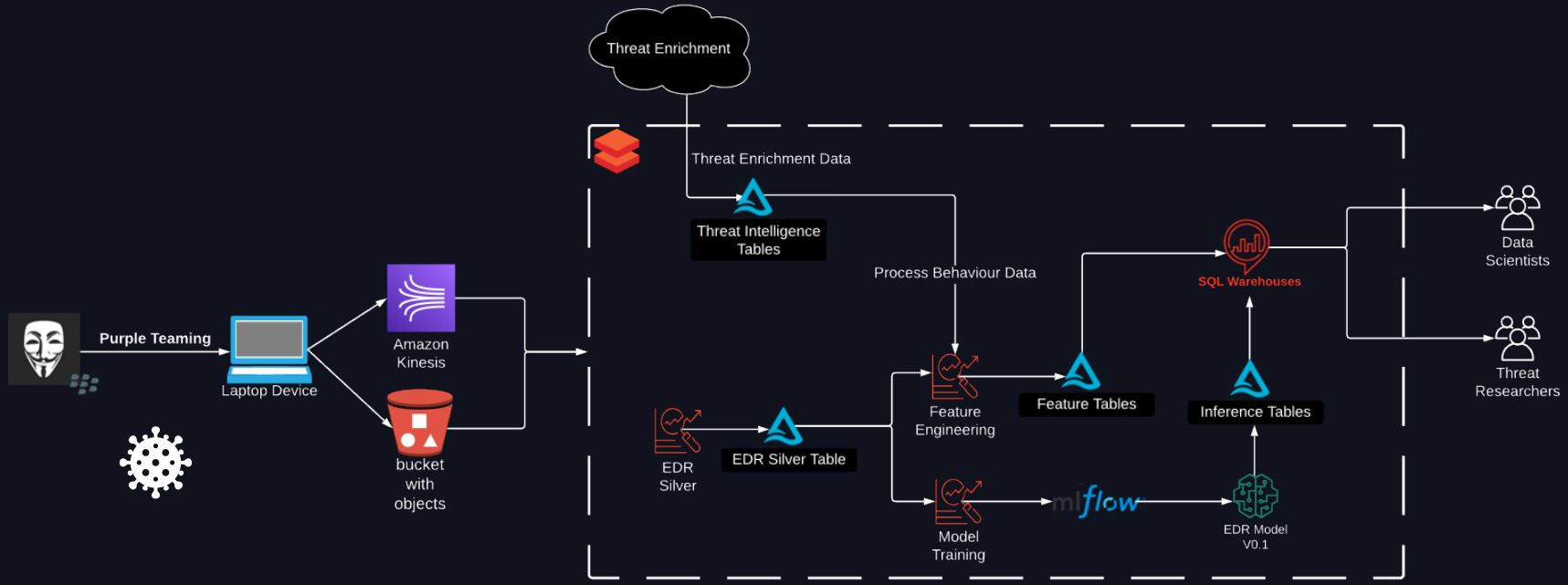
STREAMING EDR MODEL INFERENCE

Putting Telemetry to work

- EDR Telemetry data is noisy
- Over 5 million events a day per tenant
- Model Training with Apache Spark and MLFlow
- Collaborated with Threat Research

STREAMING EDR MODEL INFERENCE

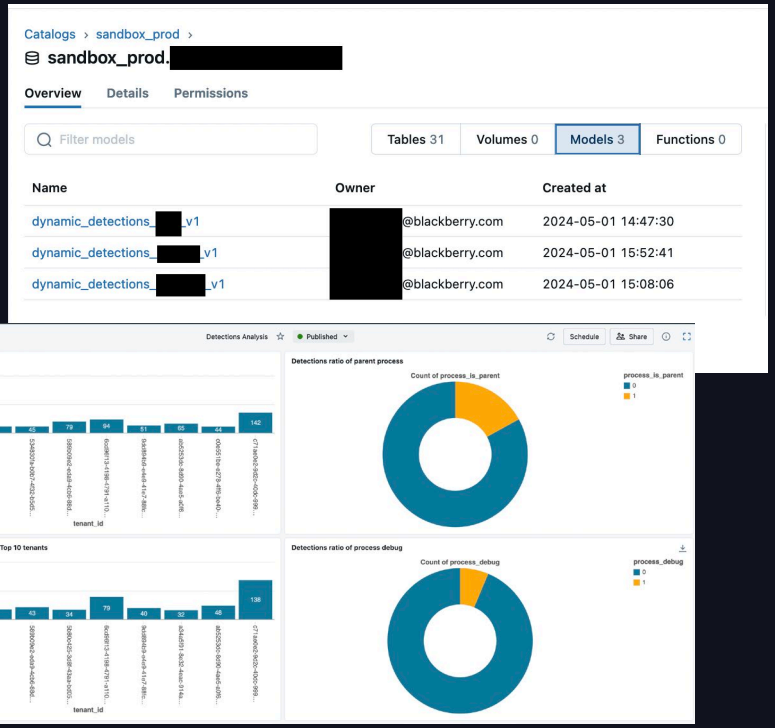
Model Training



STREAMING EDR MODEL INFERENCE

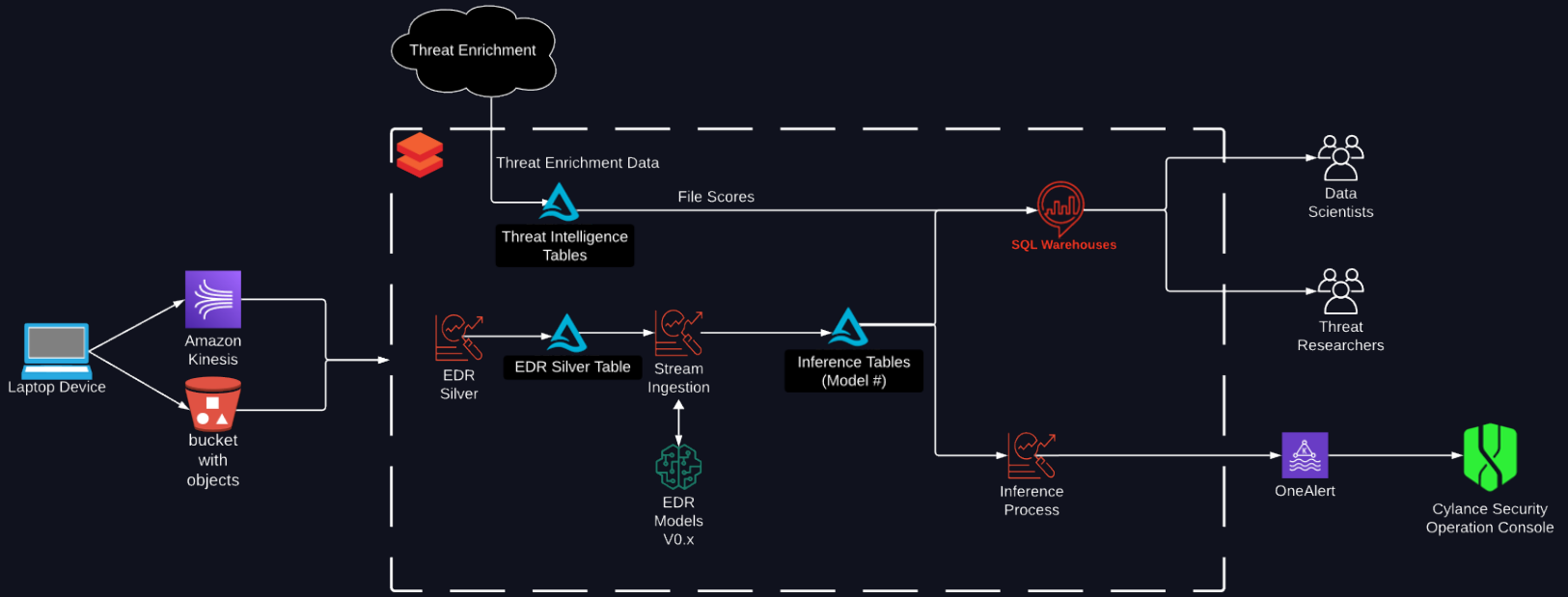
Model Training

- Wealth of BlackBerryCybersecurity Data readily available
- Common Environment enables collaboration between teams



STREAMING EDR MODEL INFERENCE

Streaming Model Pipeline



STREAMING EDR MODEL INFERENCE

Conclusion

- Dynamic Threat Modeling
- Global Threat Intelligence
- Behavioral Analysis
- Handling Large Volumes of Data

ALERT PRIORITIZATION

ALERT PRIORITIZATION

Problem

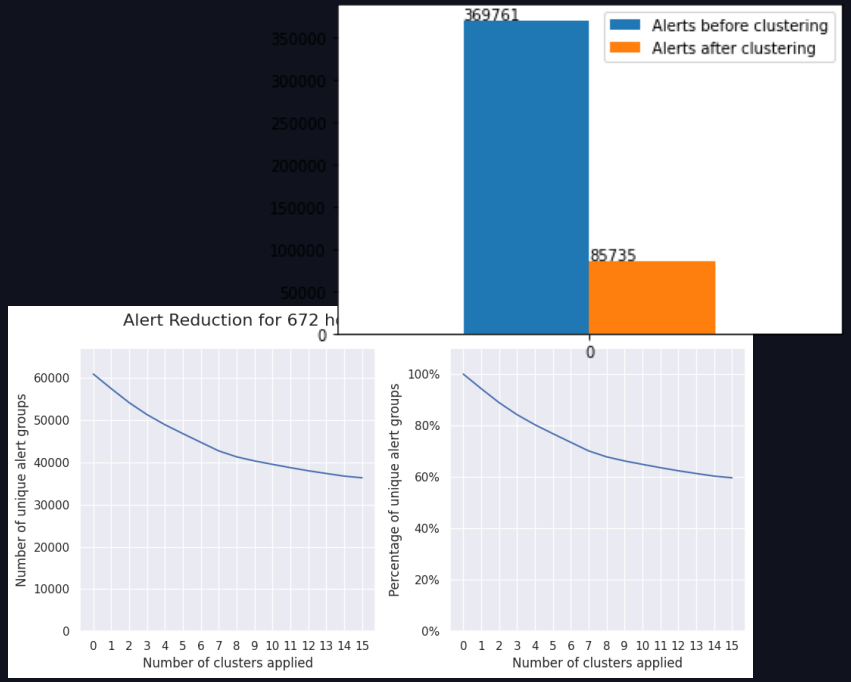
- OneAlert uses *key indicators* (KIs) to determine if an alert is unique, or if it is another occurrence of a previous alert.
- If the Kis are similar, they are counted as unique alerts

Filepath	Command
c: \windows\microsoft.net\framework64\v4.0.30319\csc.exe	"C: \Windows\Microsoft.NET\Framework64\v4.0.30319\csc.exe" /noconfig /fullpaths @"C: \Windows\TEMP\3ecp1ys3\3ecp1ys3.cmdline"
c: \windows\microsoft.net\framework64\v4.0.30319\csc.exe	"C: \Windows\Microsoft.NET\Framework64\v4.0.30319\csc.exe" /noconfig /fullpaths @"C: \Windows\TEMP\mnnh5vg3\mnnh5vg3.cmdline"

ALERT PRIORITIZATION

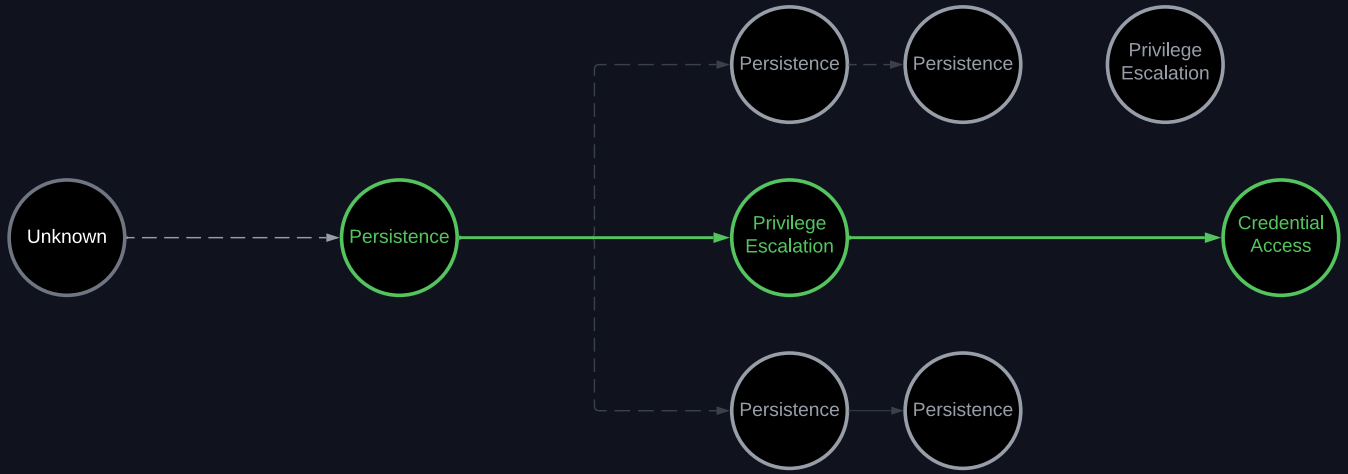
Solution

- Group similar Alerts together
- Individually Customized
- Leverage DBSCAN
- Tag Clusters by score
- Generate grouping



ALERT PRIORITIZATION

Future



- Explore other methods for prioritize and group Alerts
- Enhance Integration of ML Pipeline with Production Infrastructure
- Continue global roll out of Alert Prioritization

THE FUTURE

Data Lake to Lake House to Intelligence Platform

CYLANCE AI FOR BLACKBERRY

Supporting all Products and ecosystems

- Leverage Vector Search Index capabilities with Databricks to serve up data for LLM RAG workflows
- Data is ingested, transformed, and loaded within Databricks
- Existing Data is customized to be optimally served by Vector Search Index to ensure products such as Cylance AI has the most accurate and up to date information

THE FUTURE

Where we are going? Are we there yet?

“Create a BlackBerry Cybersecurity Data Intelligence platform to protect our customers from Cybersecurity threats”

- Expand Machine Learning and AI uses with BlackBerry Cybersecurity data
- Continue to incorporate AI into everything we do

Special thanks to Databricks team
Digan Parikh, Tal Flatt, and Amisha Singh

DATA+AI SUMMIT

Special thanks to BlackBerrys team Laura Greaves, Rejish Cheruvatta, and Helbert Diaz

DATA+AI SUMMIT

DATA+AI SUMMIT
BY databricks

THANK YOU!

Find out more about Cylance AI!