Versatility 2024

Harnessing AI for Threat Defense: Innovations in Malware Detection and DLP

VERSA

Agenda

- Pain points in cybersecurity.
- Key components.
- AI/ML for Advanced Threat Prevention (ATP).
- AI/ML for Data Loss Prevention (DLP).
- AIOps for cybersecurity.
- Questions.

Versatility 2024

VERSA

What are Today's Pain Points?

Usage of Public cloud hosted, and SaaS applications has resulted into –	***Exponential growth of complexity and risk associated with painful integration of point products*** ***Security and networking teams are faced with a constant barrage of threats and outages as the attach surface grows.***
Network and Security events are tremendously expensive –	***the cost per breach is over $4M, the cost per network outage is up to $300K per hour. <u>The cost to the brand is immeasurable.</u>***
Multiple point products have direct business impact –	***On average there are 76 different security solutions which creates problem of interworking the disparate security solutions, cost and complexity*** ***83% of Data breaches attributed to human error and 43% to misconfigurations.*** ***CISOs, CIOs and their understaffed teams are overwhelmed by massive volumes of telemetry data from point products, limited budgets, and legacy processes that do not scale***
Attackers have recognized the power of AI -	***Tools like FraudGPT or WormGPT or other uses of AI will become the next major tool used by attackers to launch sophisticated and faster attacks.*** ***Average time to respond and remediate attacks is 6 days! This needs to be reduced to hours. Innovative approaches are required for swift response.*** ***Need AI/ML intervention to level the playing field.***

The Future of Infrastructure – AI/ML Driven

Models are many. Model selection is non-trivial and varies by use case. However, Data is the key.

The Versa Platform integrates a panoramic data set from across the entire SASE infrastructure and from the WAN Edge, Cloud, Campus, remote locations, users and devices – into a unified data lake.

VersaAI™ taps into this data lake to extract AI/ML insights that are seamlessly applied across the Versa platform.

Versa AI for file/stream-based Security – Key Components

Threat Protection

AI/ML Based Malware Detection

What is it?

Multi-stage AI/ML for real-time processing of files to identify malware

Key benefits

- ✓ *Eliminates zero-day attacks covering **90 %** of file types used to transmit malware while reducing sandbox load by **75%***

Data Protection

AI/ML Based Data Loss Prevention

For dynamic, adaptive protection from data loss (unintentional or malicious)

- ✓ *Detection of sensitive data in text documents and image – enabling more accurate and dynamic data security.*

AI/ML – Advanced Threat Prevention (ATP)

Versatility 2024

VERSA

AI/ML based Malware Detection – salient features

AI/ML-based malware detection for various file-types:

PE (Portable Executable)

PDF

Docx, Pptx, Xlsx, RTF

HTML/JS (JavaScript).

Distinct ML models trained for each file-type:

Proprietary Information

Reduce reliance on static signatures:

- AI/ML-based malware detection systems can identify patterns and characteristics common to malware behaviors.

- Analyzing the way malware interacts with the system, such as how it replicates, encrypts data, or communicates with external servers, identify the underlying malicious intent of the program

AI/ML microservice – Integral part of ATP

On receiving a request for file analysis from a) on-prem device, b) SASE gateway or c) API data protection service, the ML model corresponding to the specific file type is launched.

Performs inference and returns results such as Clean, Suspicious, Malicious, with appropriate metadata.

Depending on the result and confidence score one of the following actions:

- Subsequent multi-vendor AV or sandboxing is triggered within ATP.
- The result is returned instantly, and rest of the analysis is skipped.

TRADITIONAL MALWARE DETECTION

- Reactive Detection

  Traditional methods run regex/pcre on known signatures, yara rules etc. to identify malware which are error prone.

- Static Analysis

  Examine and evaluate the code of a software program without executing it

  Signature based, Heuristic analysis, File hashing, Resource analysis, Static code analysis tools

- "Point in Time" Remediation

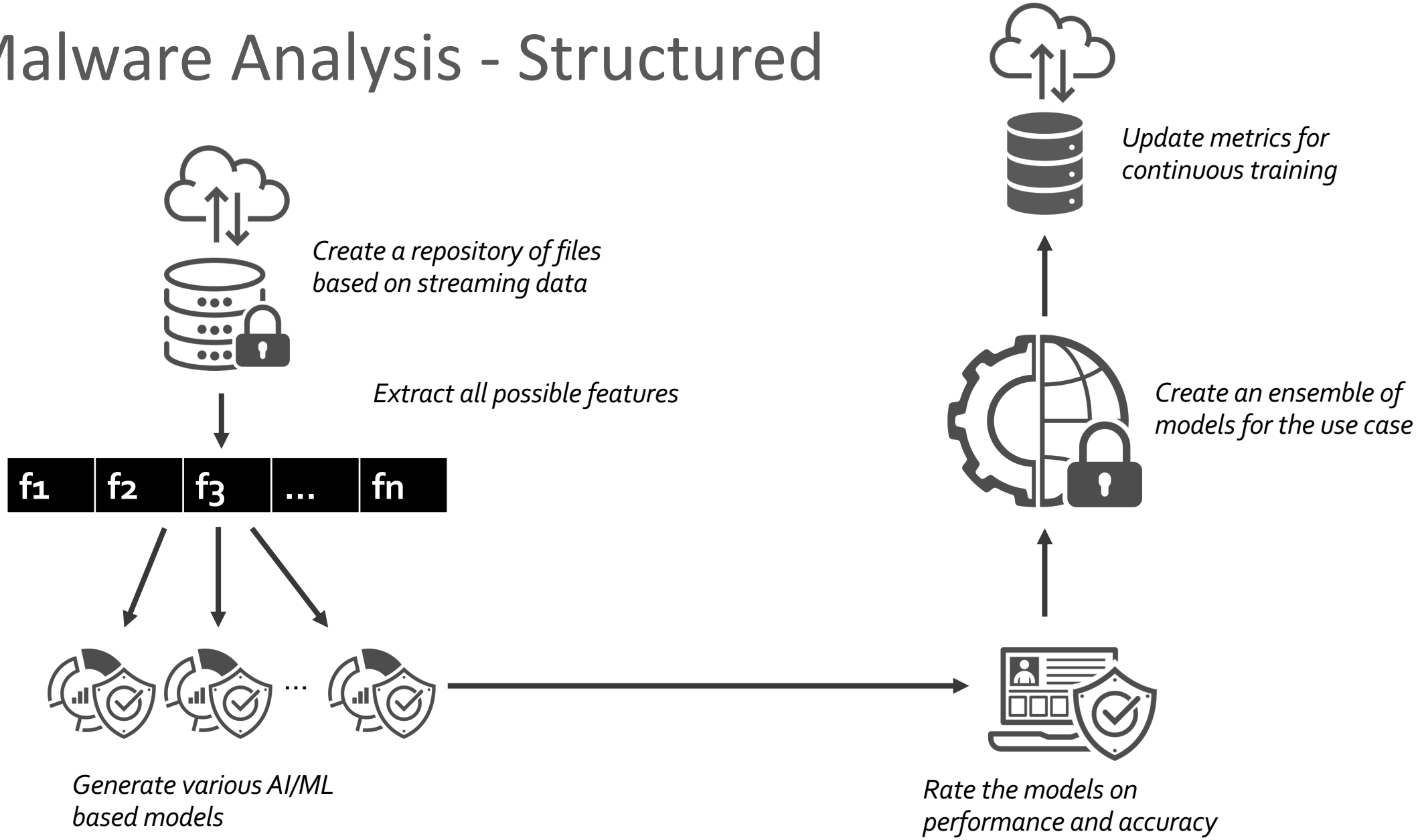
  Detection and remediation based on known attacks at a specific point of time (e.g. signatures) which may become outdated

  Traditional Malware Detection: Unable to detect Zero-Day Exploits, Polymorphic and Metamorphic Malware and Advanced Persistent Threats (APT) Identification

VERSA AI/ML BASED MALWARE DETECTION

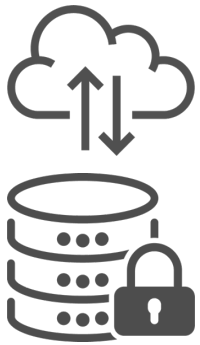
- Proactive Detection
  - ✓ *AI can identify threats based on behavior, catching zero-day and polymorphic malware.*
  - ✓ *Eliminates zero-day attacks covering 90% of file used to transmit malware while reducing sandbox load by 75% (e.g PE, PDF, docx, pptx, xlsx, JS etc.)*
- Continuous Learning & Real Time Analysis
  - ✓ *Adapts over time to new threats from updated data and models, improving its detection capabilities for zero-day.*
  - ✓ *Ensemble of Models: Trained gradient-boosted trees, Deep Neural Networks as well as fine-tuned transformer models for the different file-types.*
- Sophisticated Threat Identification & Remediation
  - ✓ *"Airtight" processes for ingesting file from any source (branch, users or SASE GWs), analysis & remediation using Advanced Threat Prevention (ATP) cloud*
  - ✓ *Detection of complex attack patterns and threat vectors by analyzing correlations across diverse data sources*

Malware Analysis - Structured



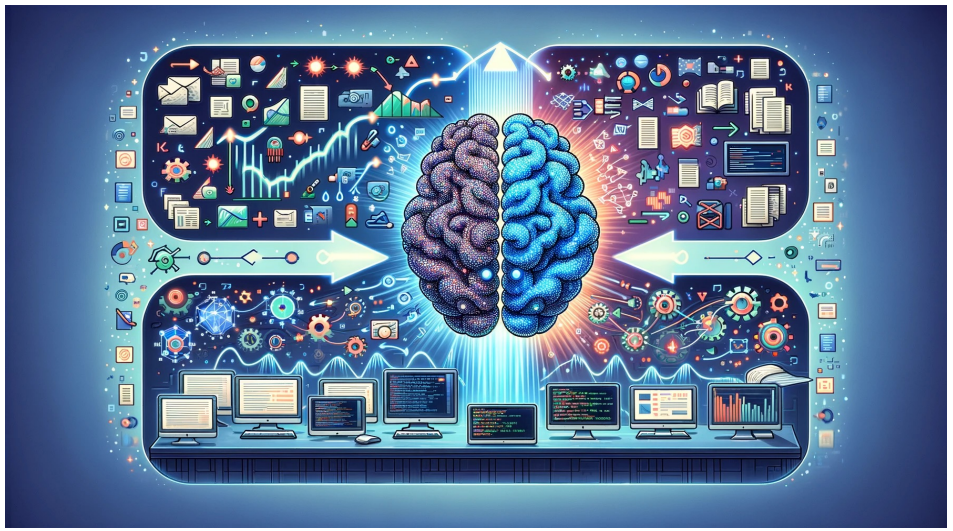
VERSA

Malware Analysis - Unstructured



Create a repository of files based on streaming data

1) *Preprocess unstructured data.*
2) *Train LLM on processed data.*



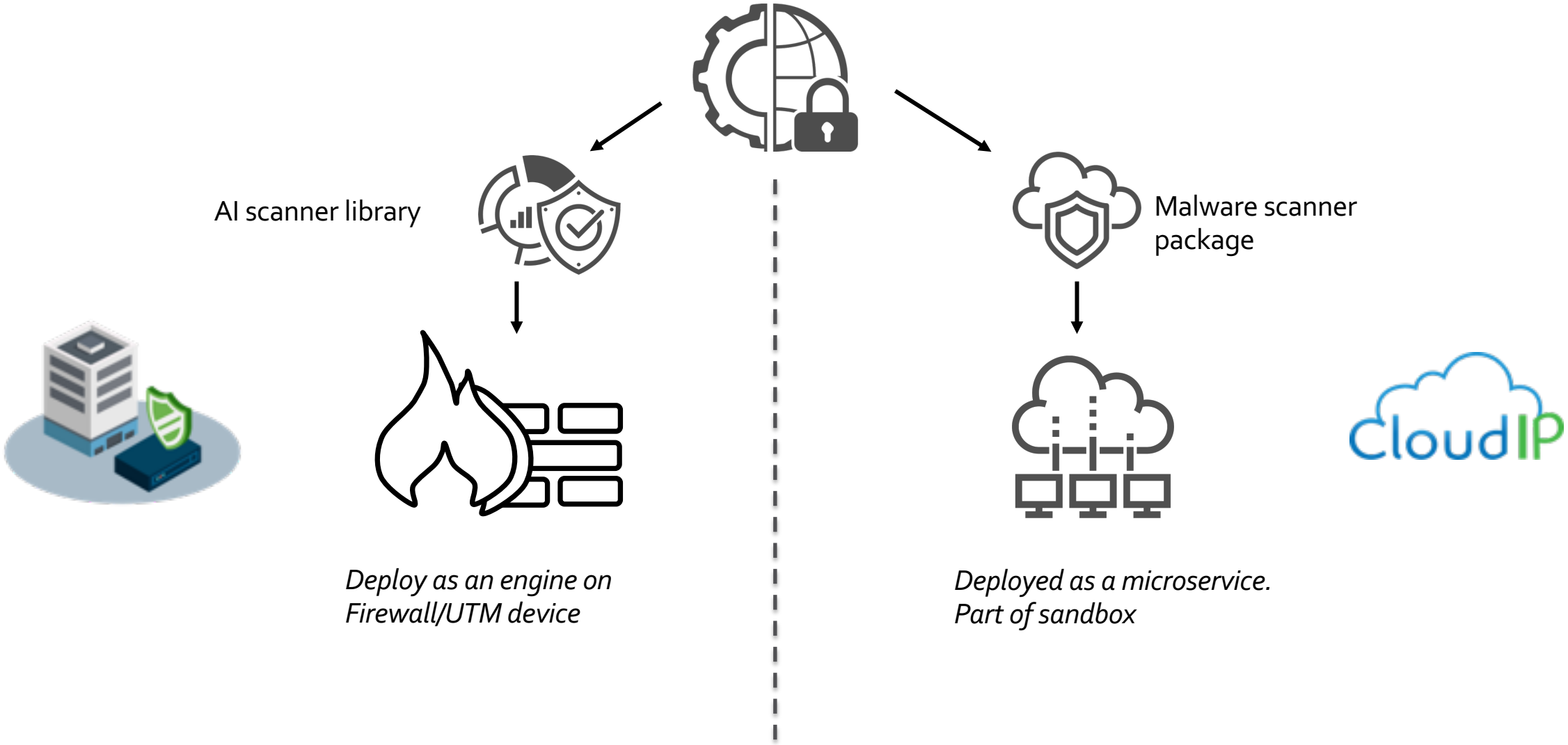
Update metrics for continuous training

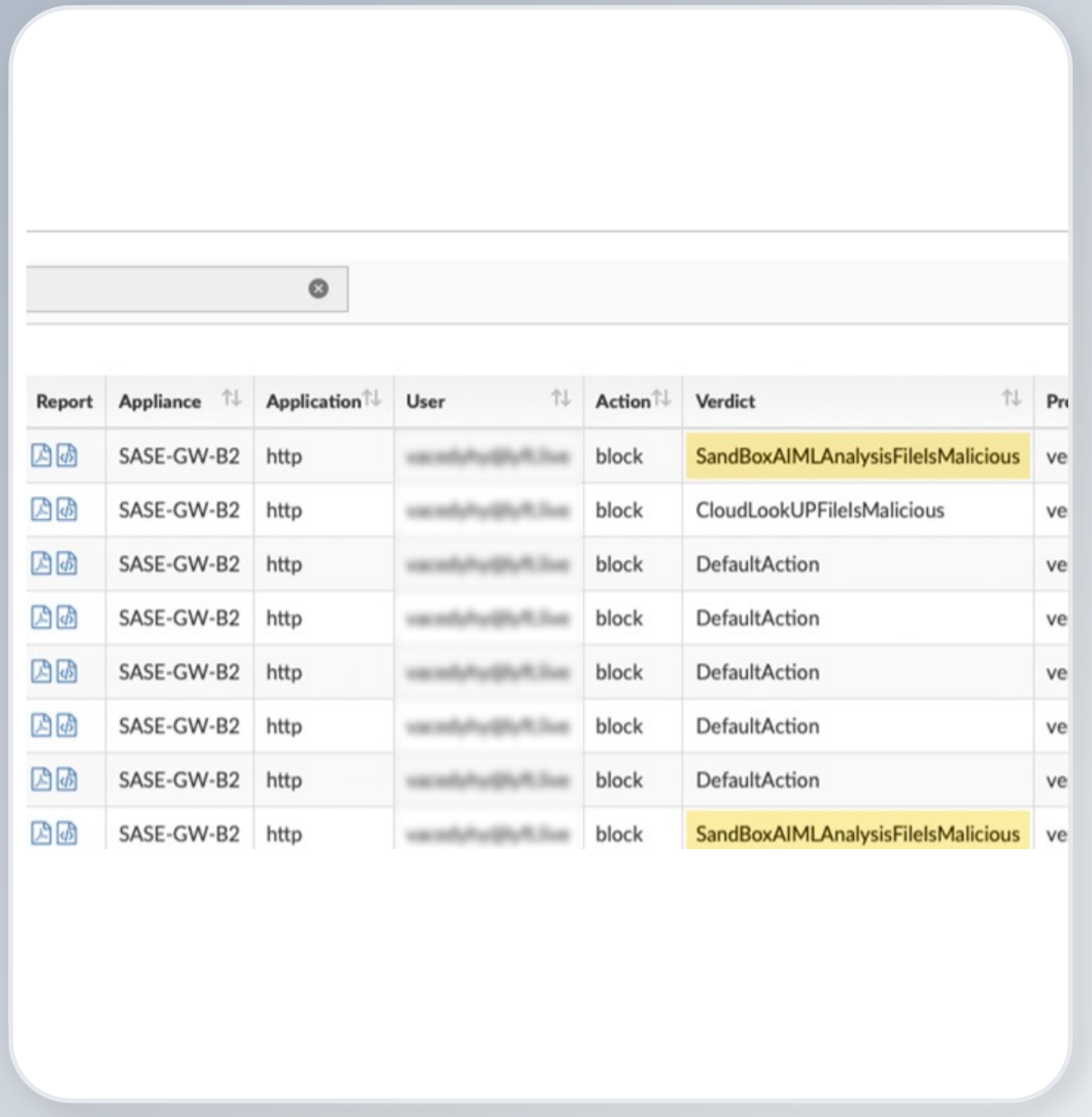
1) *Create an ensemble of models for the use case.*
2) *Optimize LLM for inference*
3) *Deploy for inference*

Rate the models on performance and accuracy

VERSA

Model Consumption Strategies





Report	Appliance	Application	User	Action	Verdict	Pr
	SASE-GW-B2	http	[redacted]	block	SandBoxAIMLAnalysisFileIsMalicious	ve
	SASE-GW-B2	http	[redacted]	block	CloudLookUPFileIsMalicious	ve
	SASE-GW-B2	http	[redacted]	block	DefaultAction	ve
	SASE-GW-B2	http	[redacted]	block	DefaultAction	ve
	SASE-GW-B2	http	[redacted]	block	DefaultAction	ve
	SASE-GW-B2	http	[redacted]	block	DefaultAction	ve
	SASE-GW-B2	http	[redacted]	block	DefaultAction	ve
	SASE-GW-B2	http	[redacted]	block	SandBoxAIMLAnalysisFileIsMalicious	ve

AI based Malware Detection

- VersaAI™ deploys multi-stage AI/ML for real-time identification of malware
- Eliminates zero-day attacks covering 90 % of file types used to transmit malware while reducing sandbox load by 75%
- Consumed as part of the sandbox as well as part of on-premise NGFW

VERSA NETWORKS

AI/ML – Data Loss Prevention

Versatility 2024

VERSA

AI/ML based DLP – high-level components

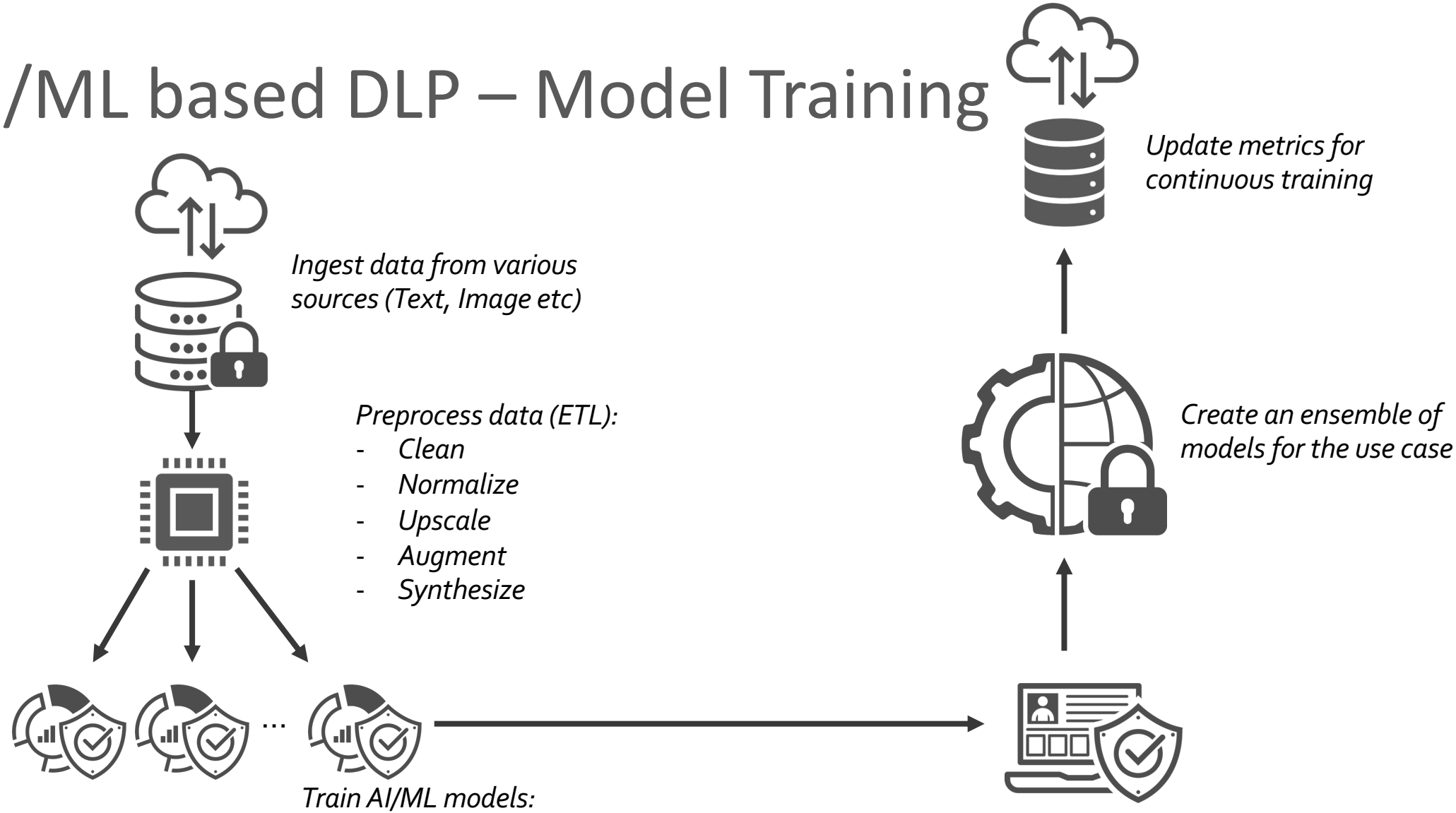
DLP for multiple document types: - *Text, PDF, DOC, Images etc.*

Image and Text pre-processing for DLP training and inference

Bleeding edge transformer models used to detect the type of document (e.g. Credit cards, Passport, source-code,) and extract multi-modal content.

Fine-tuned LLMs for classification and detection of sensitive/PII data.

AI/ML based DLP – Model Training



Ingest data from various sources (Text, Image etc)

Preprocess data (ETL):
- *Clean*
- *Normalize*
- *Upscale*
- *Augment*
- *Synthesize*

Train AI/ML models:
- *Embeddings models*
- *CNN based classifiers.*
- *NER models*
- *Transformer/LLM based models*

Rate the models on performance and accuracy

Create an ensemble of models for the use case

Update metrics for continuous training

Traditional DLP

- No “context” in analysis

  Traditional methods rely on pre-defined pattern matches which may lead to false positives or negatives in certain attack scenarios

- Static Detection Techniques

  a) Keyword Matching (words such as "confidential) b) Regular Expression (Regex) Matching for Pattern Recognition (Social Security #s) c) Document Fingerprinting (unique hash or "fingerprint") d) File Type and Metadata Analysis (author, creation date, file size) e) Data Classification and Tagging (classify documents by tagging). Pattern matching ”rules” that can be circumvented easily by learning the detection techniques with malicious AI

- Limited Adaptation

  Static analysis without real time adaptation of ever-changing threat vectors

VERSA AI/ML BASED DLP

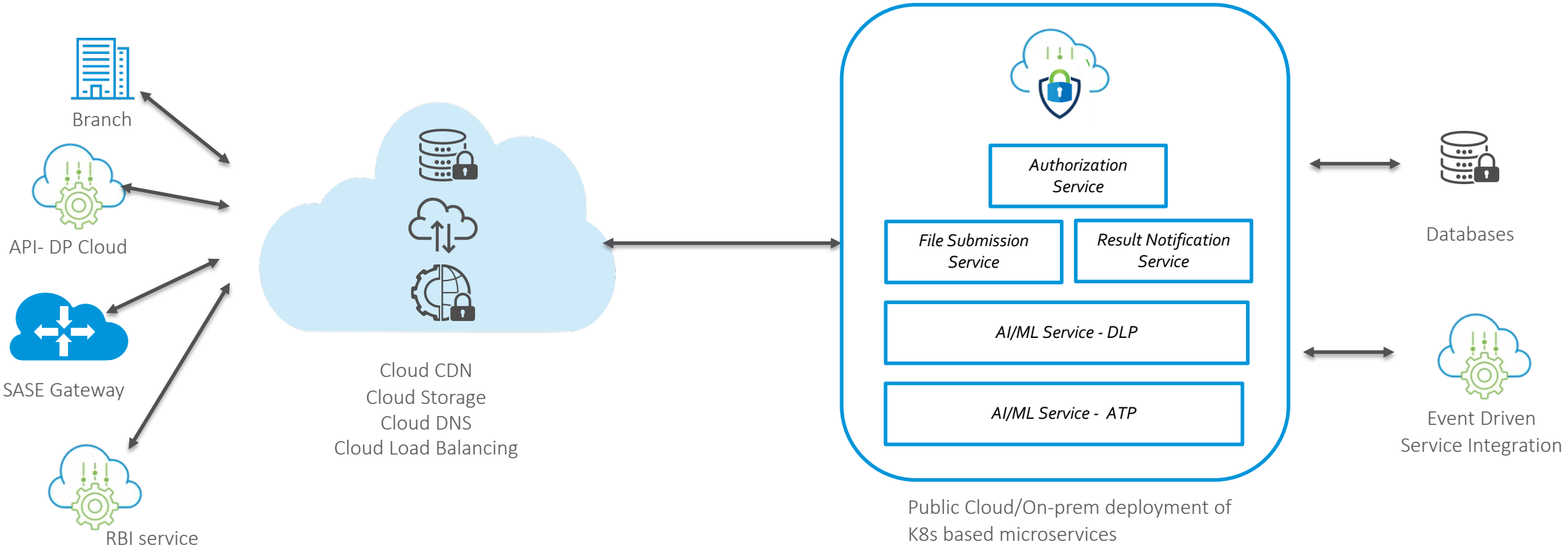
- Enhanced Detection with Contextual Analysis
  - ✓ *AI algorithms analyze behaviors and patterns to accurately distinguish between normal activities and potential data breaches, reducing false positives.*
  - ✓ *Understands the context (spatial or temporal) of data usage, allowing for more nuanced protection strategies that adapt to different scenarios.*
- Dynamic Adaptation with Real Time Protection
  - ✓ *Continuously learns from new data, improving its understanding of what constitutes sensitive information and potential threats over time*
  - ✓ *Monitors data in motion, at rest, and in use in real-time, providing immediate response to potential data leaks or unauthorized access*
- Advantage AI/ML
  - ✓ *Anomaly Detection in Data Movement*
  - ✓ *Detection of Sensitive Information in Unstructured Data*
  - ✓ *Adaptive Protection Against Evolving Data Types and Policies*

AIOps for Cybersecurity

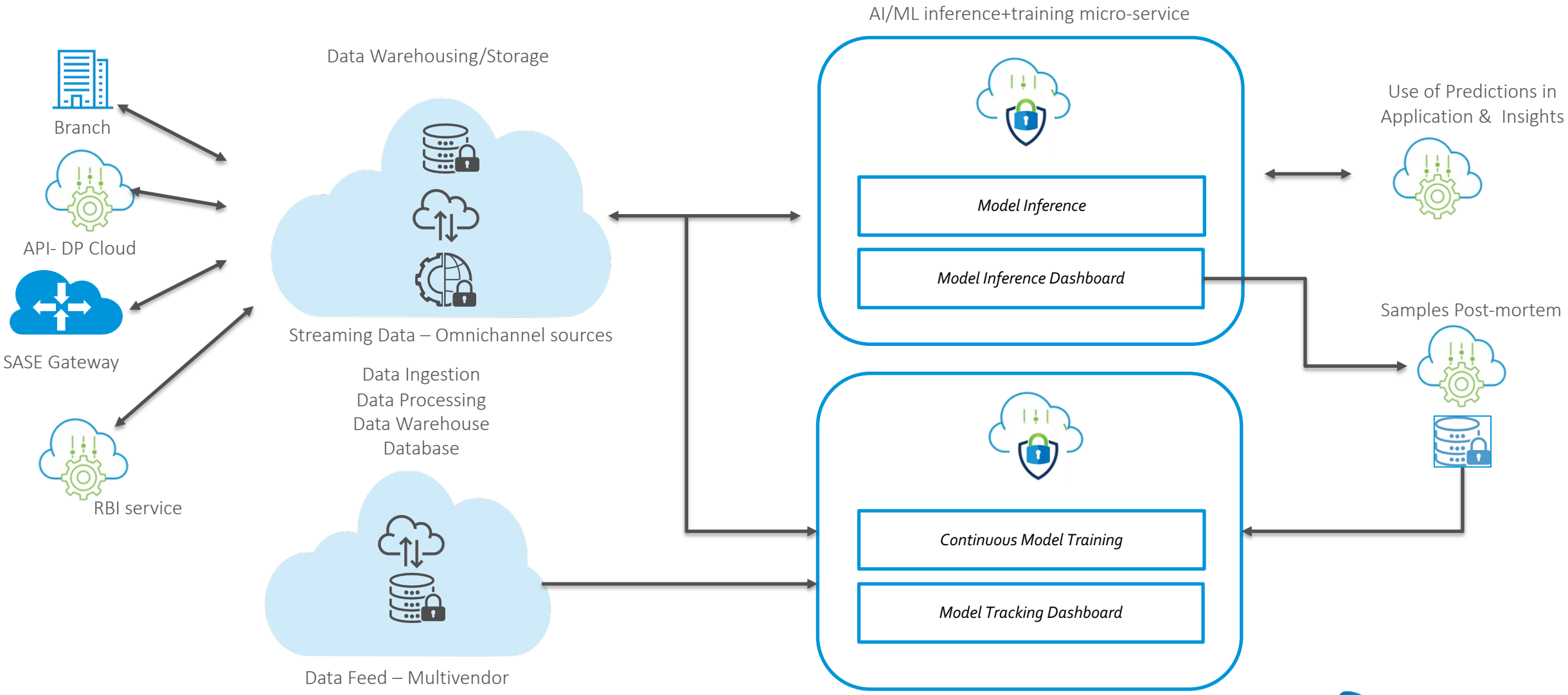
Versatility 2024

VERSA

Public Cloud Based Versa AI/ML Services Flow Diagram



Public Cloud Based Versa AI/ML ATP/DLP Data Pipeline



VERSA

Questions

VERSA

Thank you

VERSA