Universal Data Transformation Fabric

How AI-Native Infrastructure works for OCSF Mapping

Speakers

Yichen
Fleaker

Paul Agbabian
Splunker
OCSF Architect

Gavriel
Meir-Levi
Fleaker

Agenda

1. Introducing Fleak
2. Challenges in analyzing enterprise security data
3. What is OCSF?
4. Where Does Fleak Operate in Production?
5. Demo: AI Native OCSF Mapping
6. Case Study

Fleak:

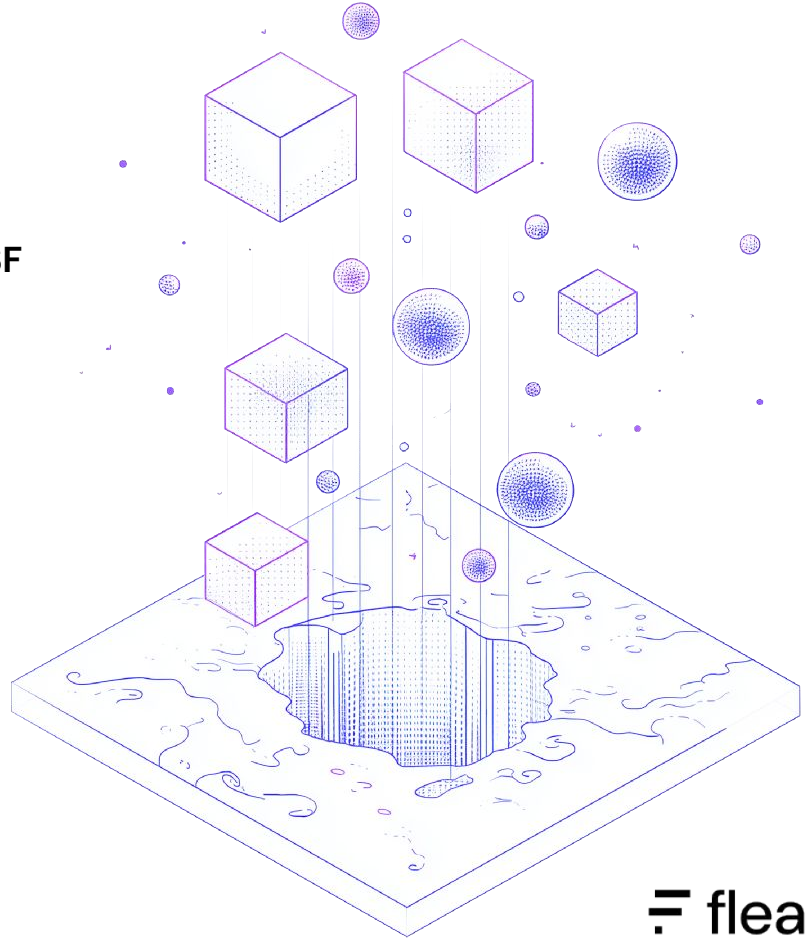
The Universal Data Translator

- Pick a destination schema framework such as **OCSF**
- Deploy on the edge or in the cloud
- Process millions of events per second
- Self-adapt to input data drifts
- Complete governance and data sovereignty
- Real-time translation

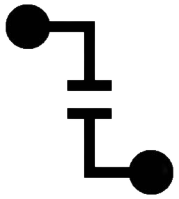
OCSF TIA CSF

CEF IEC OT

CUSTOM

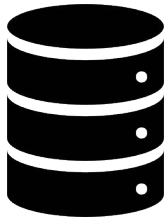


Challenges in analyzing enterprise security data



Inconsistent and incomplete data

Logs and alerts in varying formats scattered across the organization in tough-to-find data silos



Growing volumes of security tools and data

SOC teams often manage 45+ security tools — integration is complex

Inefficient use of data across use cases

Security teams lose time mapping data instead of detecting threats

Multiple proprietary log schemas

Duplicate ETL pipelines for each tool.

Open Cybersecurity Schema Framework (OCSF)

An open standard that can be adopted by anyone to simplify security data normalization

Open-source project to deliver a simplified and vendor-agnostic taxonomy for security data that can be adopted in any environment, application, or solution provider

Speed up data ingestion and analysis without the time-consuming, upfront normalization tasks

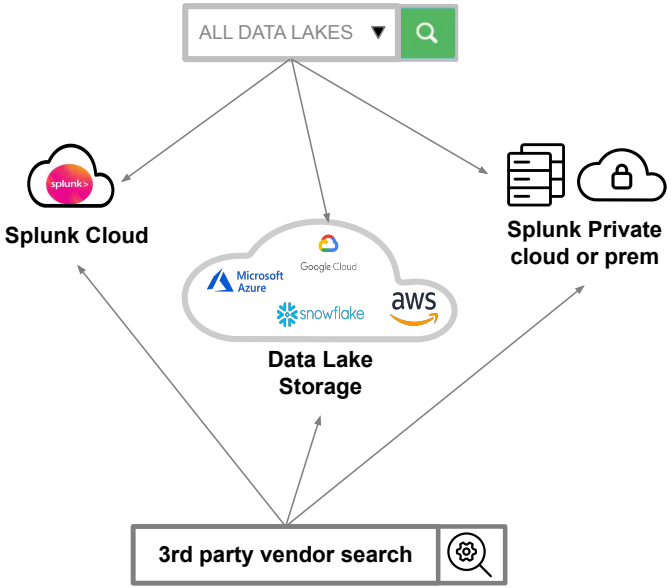
Combine data from OCSF-compliant sources to break down data silos that slow security teams

Officially joined the Linux Foundation in 2024

Industry-driven, community-based approach with over 1,100 participants across 200+ organizations

Fundamentals of the framework

Separation of schema from data model implementation for data lake compatibility



Profiles – common aspects across event classes, objects and categories

Categories – organize event classes

Classes – represent events of a type, are sets of attributes from an attribute dictionary of standard fields and objects

Objects – entities and resources

Attributes – from a dictionary

Types – constrained and validated

Normalized across products:

```
SourceIP = SRIP = src_ip
```

The schema is **open and extensible** by customers and vendors, allowing for **cross product normalization**

It is **agnostic** to implementation and storage format (e.g. JSON, Parquet)

It has a **common dictionary of attributes and objects**

Every **event class** has a **disposition** or outcome

The **type_uid** combines the event_class_uid and the activity_id and **gives the event 'meaning'**

How OCSF Works

OCSF simplifies security data integration through a standardized schema:

Event producers

- Security tools
- Cloud services
- Applications
- Network devices

Data consumers

- SIEM
- Data Lakes
- Analytic Tools
- AI/ML Models

Key benefits include:

- One-time mapping for producers
- No custom parsers needed for consumers
- Hybrid cloud & multi-vendor ready
- AI/ML-ready data format

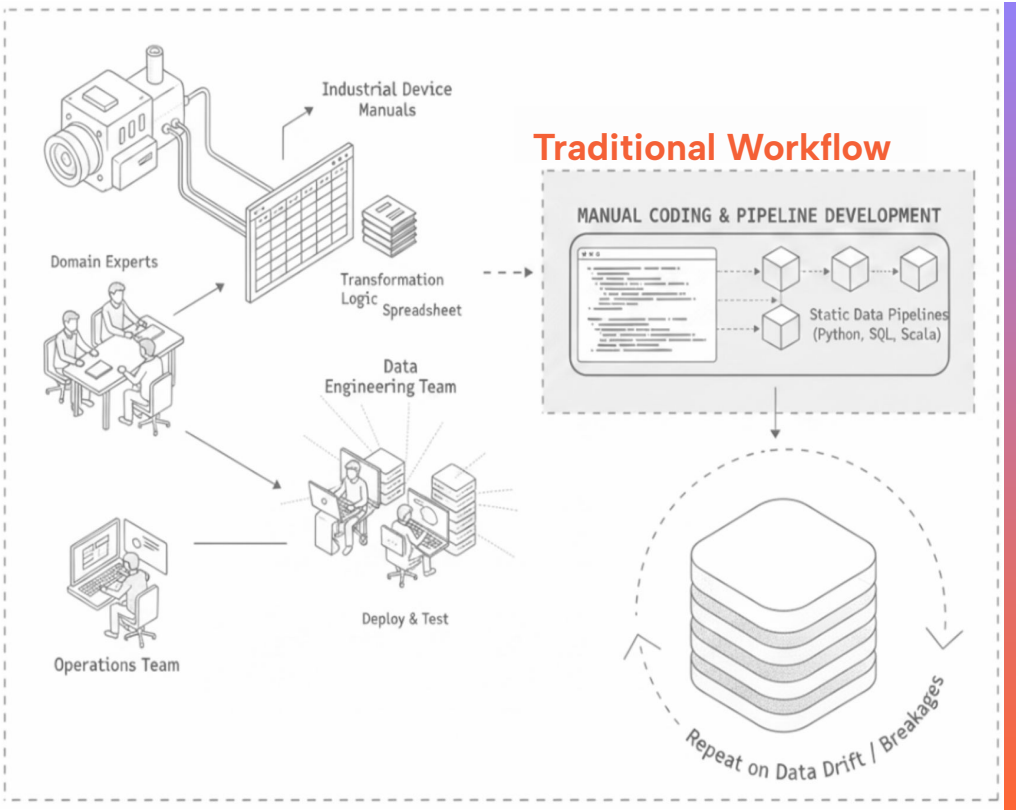
What This Means For The US Military

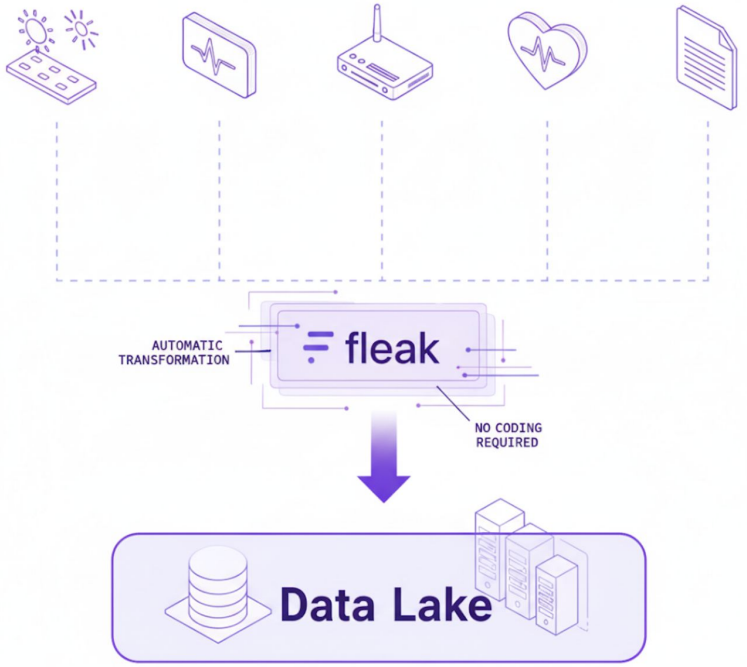
The Intelligence Cycle: From Data to Decision

The Intelligence Cycle Must:

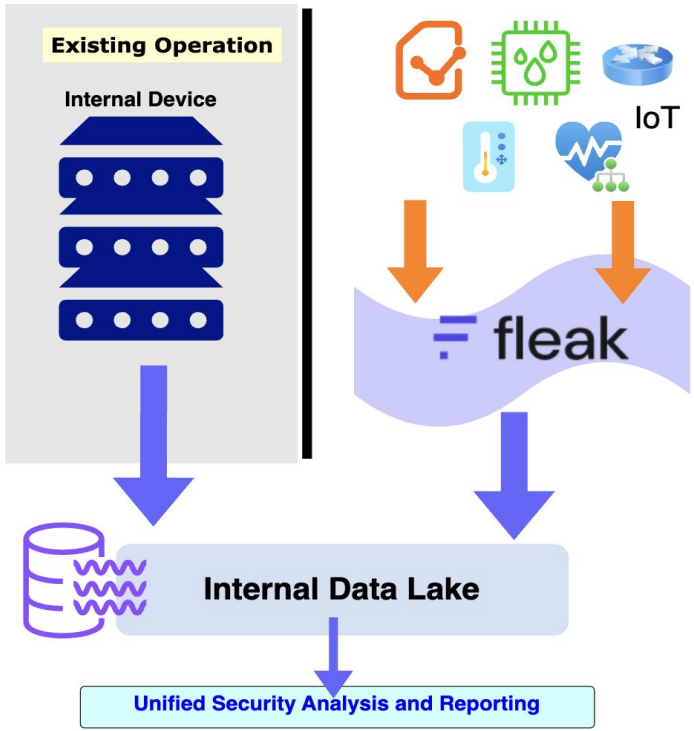
- *Enable AI at scale*
- *Produce results in real-time*
- *Accommodate thousands of data sources and formats from legacy past to cutting-edge future*
- *Adapt and self-heal*

Traditional Data Flows VS AI Native Data Fabric





Case Study: Driving Efficiency and Scale for a F500 Enterprise



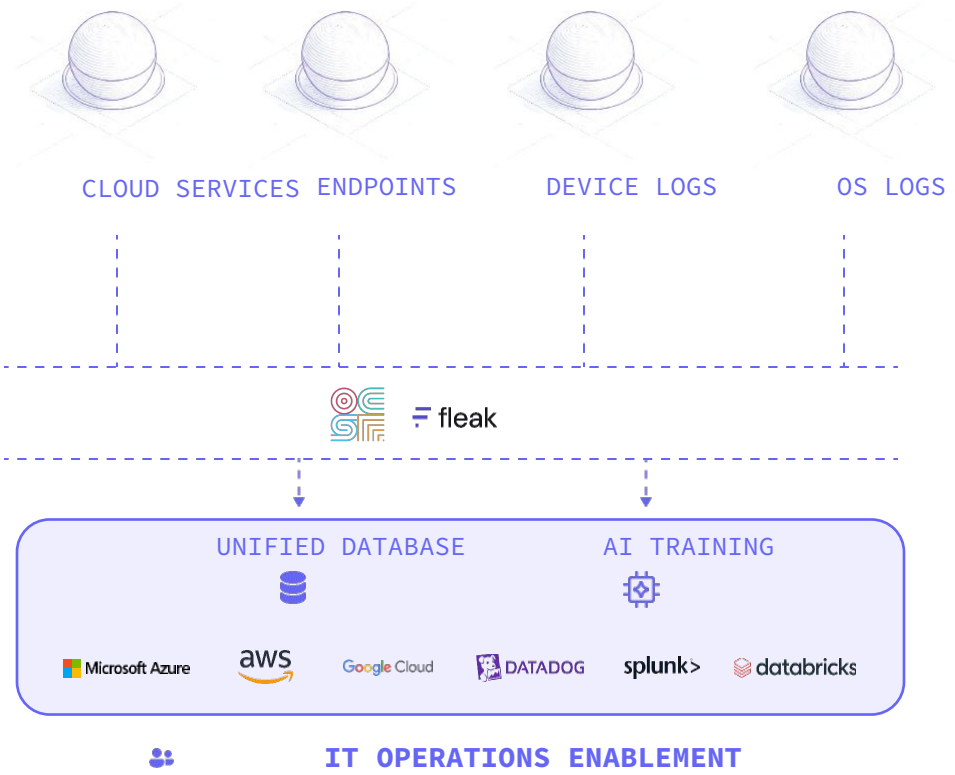
⏰ 6 months/source – 1 week

💵 90% cost reduction

🛡️ 10x more integration of disparate data sources

Where Does Fleak Operate in Production

- Between legacy infrastructure (Data Producer) and modern AI workloads (Data Consumer)
- Self-evolving data fabric (**in-motion**) that learns and adapts to changing business requirements
- Zero storage at millions EPS performance

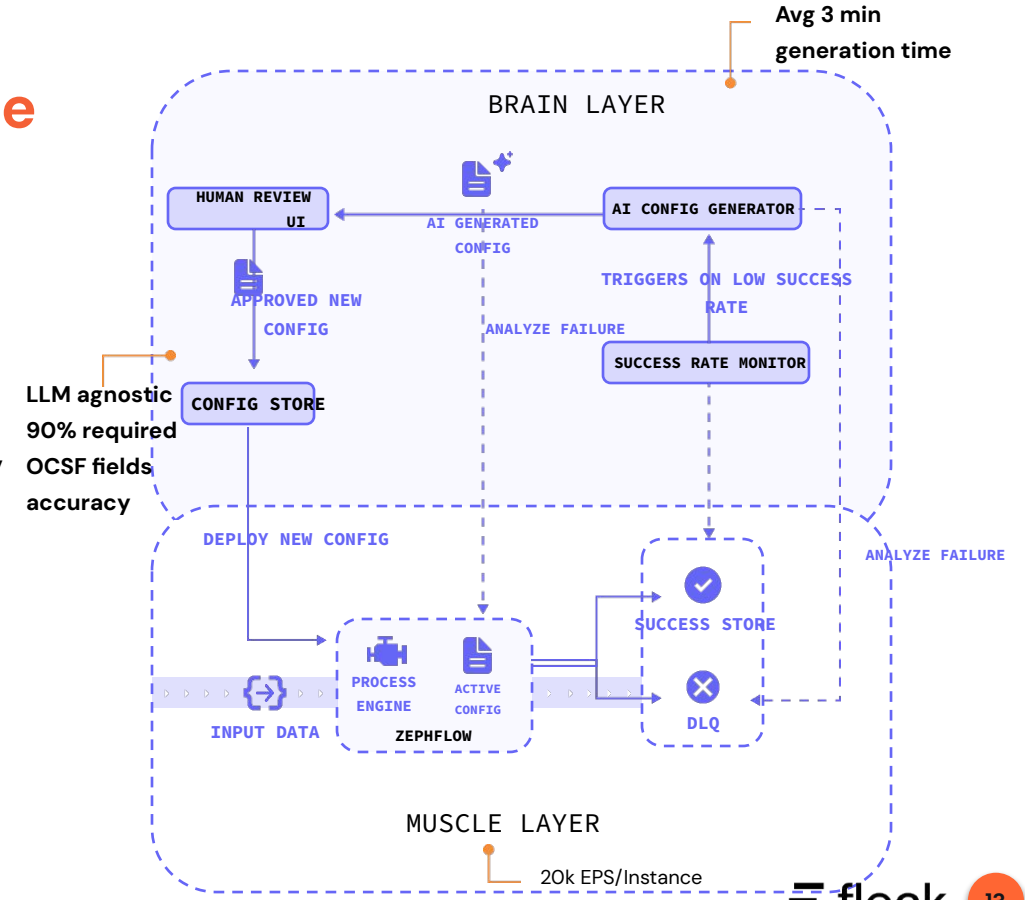


The Self-Healing Architecture

Self-Improving AI for Unmatched Precision → Every data point makes our AI smarter

Beyond Mapping Intelligence → Our brain layer can easily expands to full automation: Enrichment, Routing, and MORE...

Air Gapped Deployment → Data tagging and template generation at unprecedented speed within your own environment



demo

Proudly Partnering with

paloalto NETWORKS

ATLAS AIR WORLDWIDE

IBM

databricks

DENSO

CISCO

NetFoundry

expel

Thank You

Try AI Native OCSF Mapping at:
ocsf.ai
For Other Schema Translators, Contact us at:
contact@fleak.ai