AI.dev
Open Source GenAI & ML Summit

CASSANDRA SUMMIT

Building a Global CyberGrid on Cassandra

Charles Herring, WitFoo

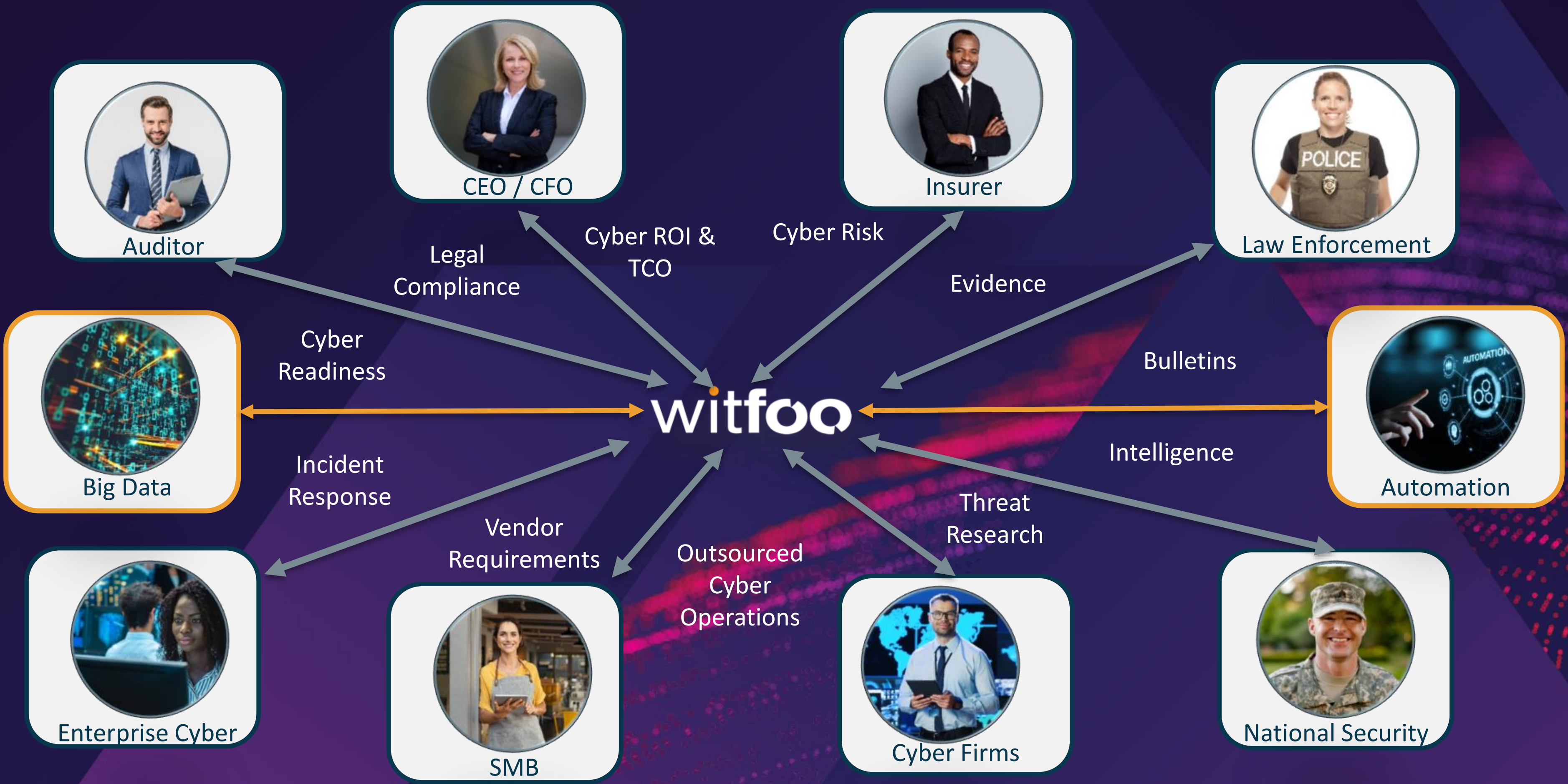
CharlesHerring.com

#AIdotDev | #Cassandra Summit

THE LINUX FOUNDATION

Agenda

- Scope of WitFoo Project
- Distributing Configuration across Clusters
- “Predestination of Data”
- Keyspace Configuration
- NLP & Graph on Cassandra
- Federating Data and Operations



witfoo

Auditor

CEO / CFO

Insurer

Law Enforcement

Big Data

Automation

Enterprise Cyber

SMB

Cyber Firms

National Security

Legal Compliance

Cyber ROI & TCO

Cyber Risk

Evidence

Cyber Readiness

Bulletins

Intelligence

Incident Response

Vendor Requirements

Outsourced Cyber Operations

Threat Research

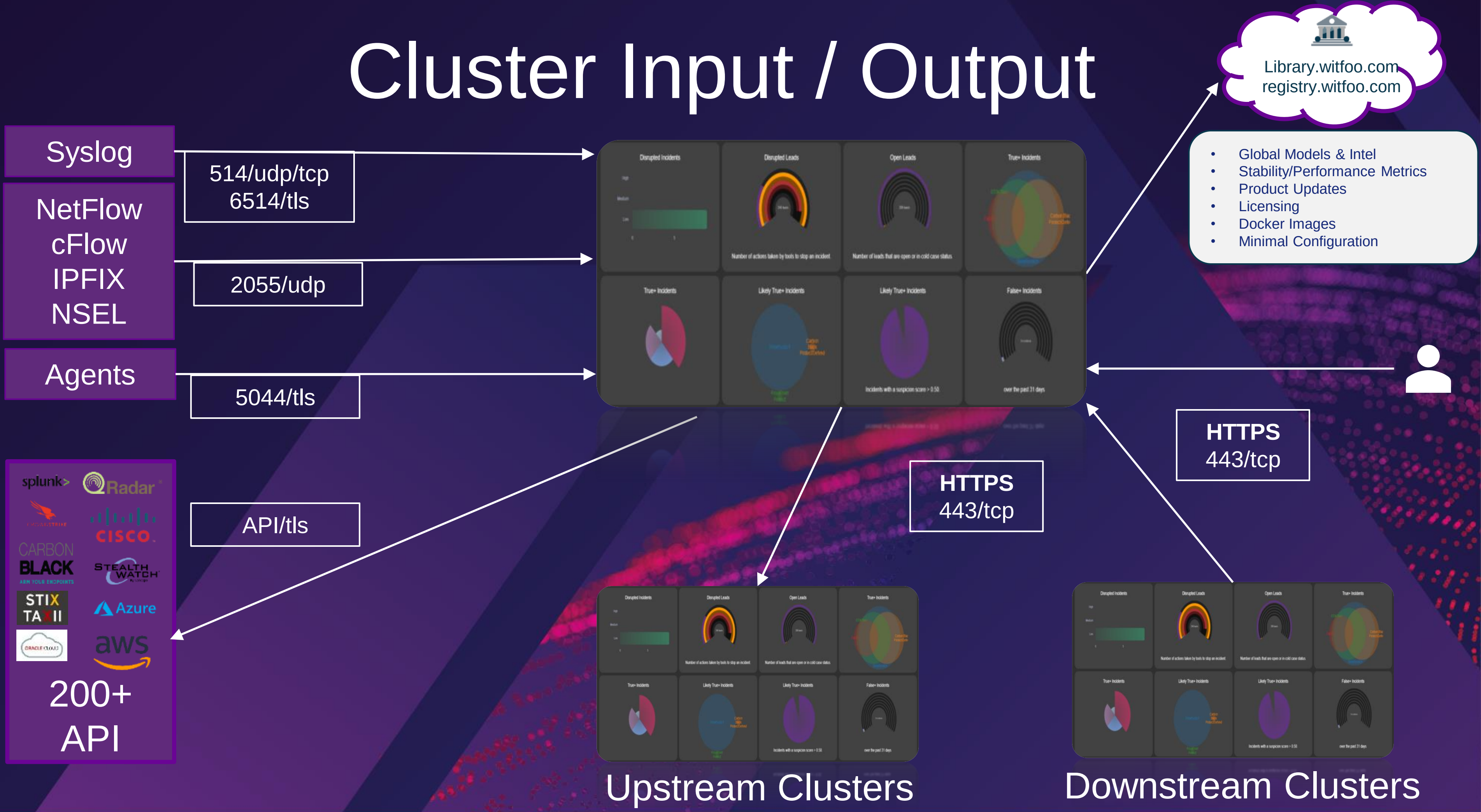
WitFoo Scope

100’S OF ORGANIZATIONS

100GB TO 10TB INGEST DAILY EACH CLUSTER

OPERATIONS ACROSS CLUSTERS

Cluster Input / Output



Library.witfoo.com
registry.witfoo.com

- Global Models & Intel
- Stability/Performance Metrics
- Product Updates
- Licensing
- Docker Images
- Minimal Configuration

Syslog

514/udp/tcp
6514/tls

NetFlow
cFlow
IPFIX
NSEL

2055/udp

Agents

5044/tls

API/tls

200+
API

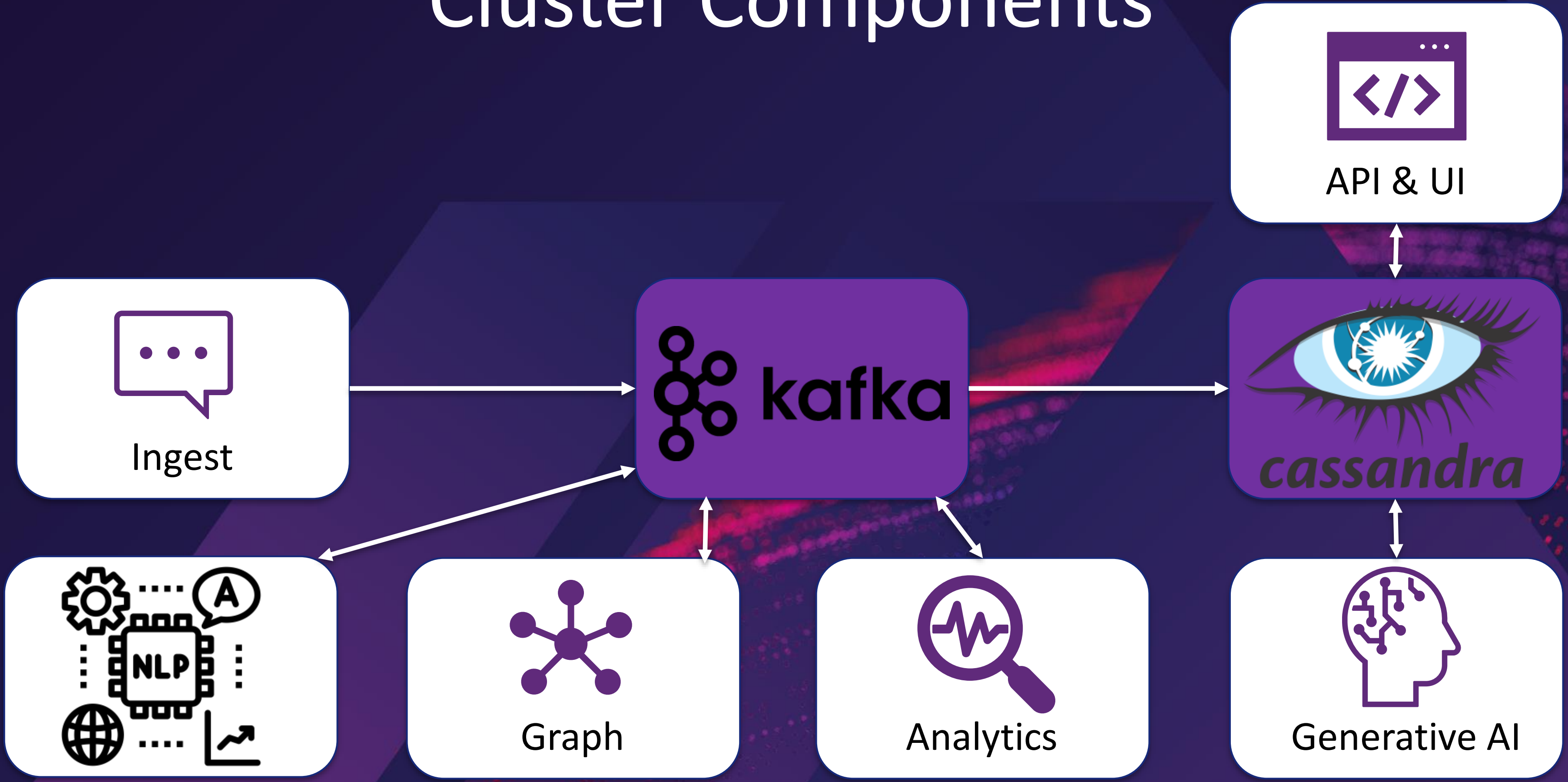
HTTPS
443/tcp

HTTPS
443/tcp

Upstream Clusters

Downstream Clusters

Cluster Components



Diverse Node Deployments

Hypervisor & OnPrem	Cloud	SaaS / Managed	Hardware & Engineering
vmware™	amazon web services™	ardalyst	accenture FEDERAL SERVICES
Hyper-V	Microsoft Azure	impelix	ardalyst
ubuntu®	Google Cloud	RIMSTORM	Raspberry Pi roqos®

Configuring Nodes

- WitFoo Agent (wfa)
  - Script inputs: DC/Rack, License Key, Roles
  - Fetches from Library: Seeds and Secrets
  - Pulls Docker Images
  - Launches Containers with config
  - Ships metrics to Library
    - (*Metrics-Driven DEVOPS* Talk on CharlesHerring.com)
- Library
  - Issues alerts to WitFoo, Cluster & Customer

Predestination of Data

The entire lifespan of a datum must be established at its birth. Comprehension of syntax, source and intent must be extracted. Inference and potential impact of the datum must be established. Nature of creation and transmission must be preserved. All expected evolutions and iterations of the data need to be established for processing. The death (TTL) of the datum must be established at persistence.

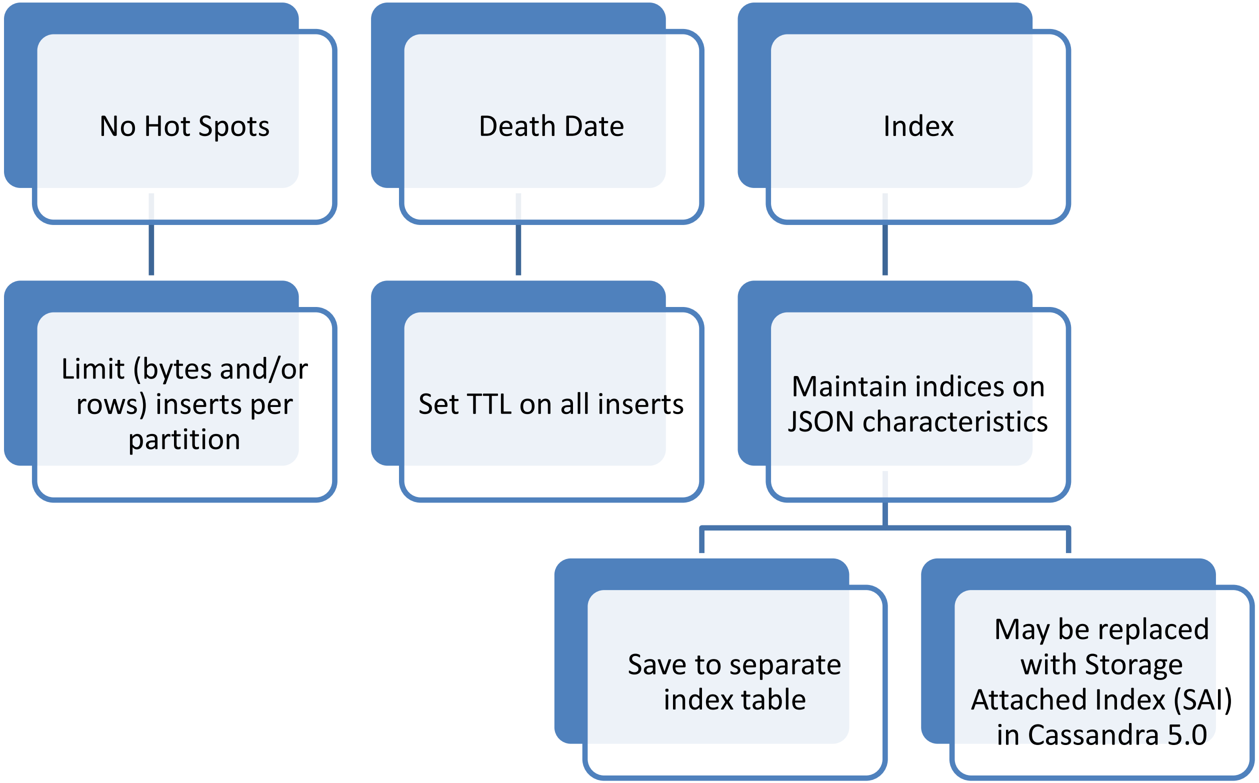
Keyspace Configuration

- Replication Factor: 3 (default)

```
(partition timeuuid, org_id
uuid, created_at timeuuid,
object text, PRIMARY KEY
((org_id, partition),
created_at)) … AND compression
= {…LZ4} AND
default_time_to_live = 0
```

- Replication
  Cassandra 101
- partition
  Limit number of rows in partition (hotspots). Gives creation time. Can be used in time-based partition scans.
- org_id
  Designate the cluster that owns data.
- (org_id, partition)
  Non-colliding, finite grouping of rows. Quick partition fetching.
- ((org_id, partition), created_at))
  Primary key with rapid single row fetching.
- object JSON Structure
- Compression
  Faster IOPS, less disk
- TTL “free” delete

Insertion Principles



No Hot Spots

Limit (bytes and/or rows) inserts per partition

Death Date

Set TTL on all inserts

Index

Maintain indices on JSON characteristics

Save to separate index table

May be replaced with Storage Attached Index (SAI) in Cassandra 5.0

Fetching Principles

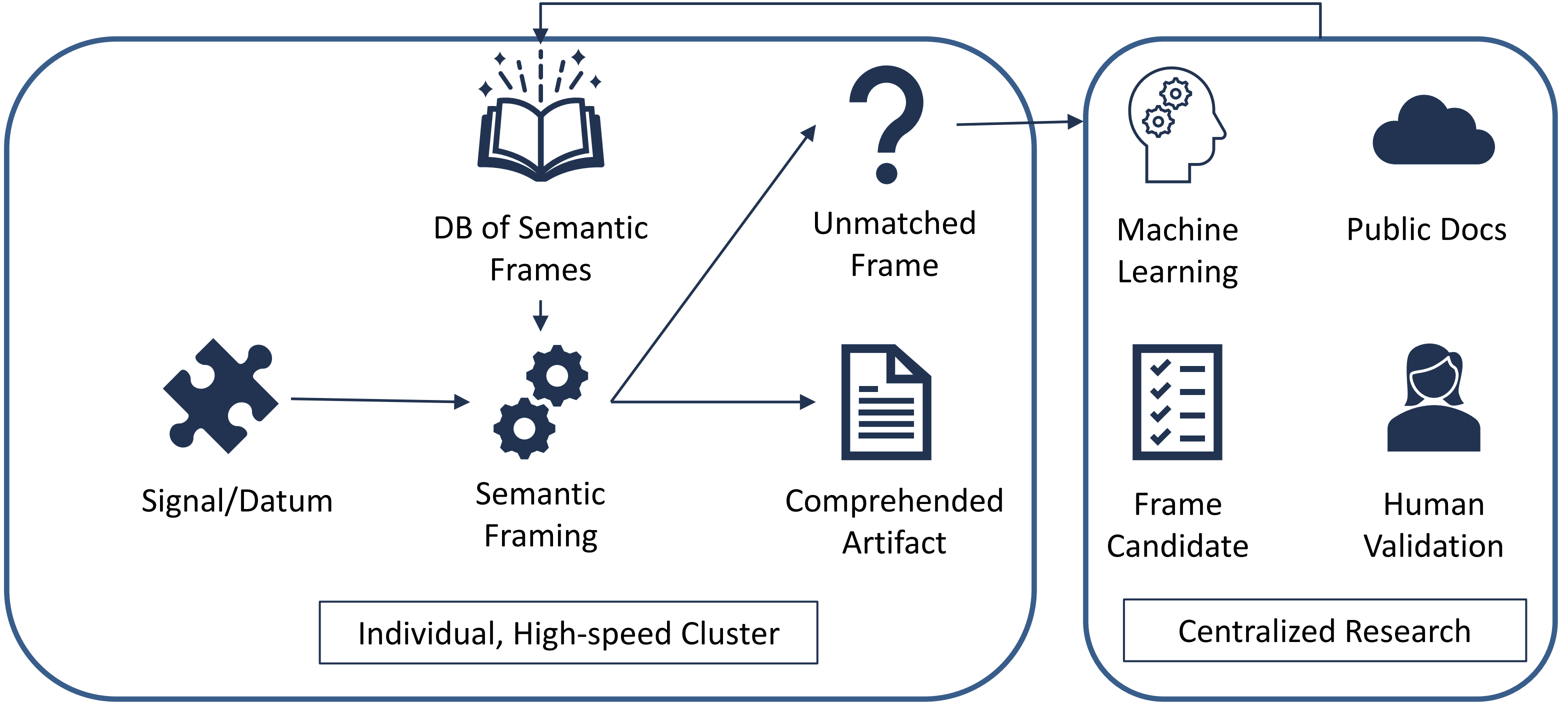
“Index Candidates”

- In time range (timeuuid)
- Match cluster (org_id)
- Object JSON index can match query

Fetch partition (org_id, partition)

Reduce results to row matches on query

Augmented Natural Language Processing



DB of Semantic Frames

Signal/Datum

Semantic Framing

Unmatched Frame

Comprehended Artifact

Individual, High-speed Cluster

Machine Learning

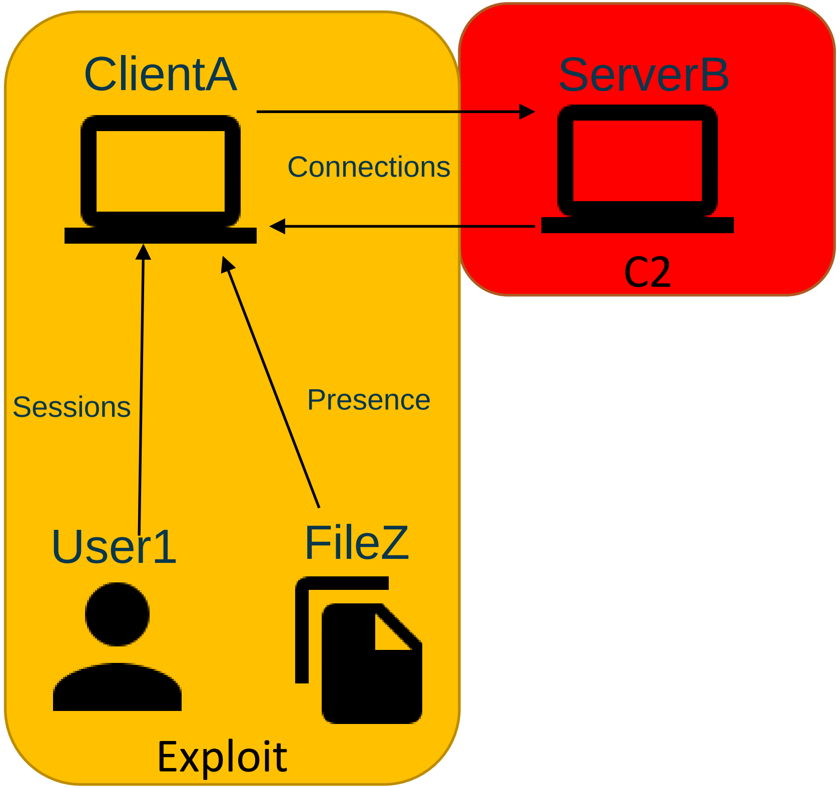
Public Docs

Frame Candidate

Human Validation

Centralized Research

Object Oriented Organization



Artifacts

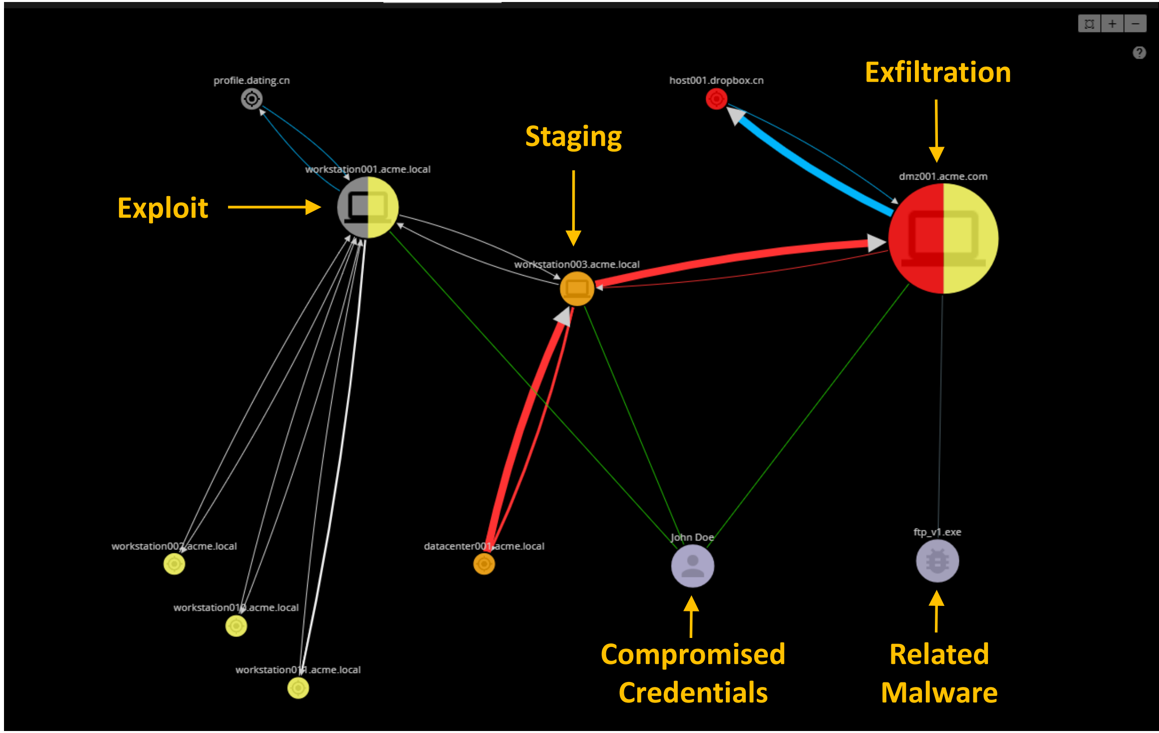
- ClientName: ClientA
- ClientIP: 10.10.10.43
- ClientMAC: 00-DC-EF-23-15-12
- Product: MS DHCP
- MessageType: DHCP Lease
- Intent: Asset Info

- ClientName: ClientA
- User: User1
- File: FileZ
- Product: Crowdstrike Falcon
- MessageType: Malware Detected
- Intent: Exploit Detection

- ClientIP: 10.10.10.43
- ServerName: ServerB
- Product: Cisco Firepower
- MessageType: C2 Detected
- Intent: C2 Detection

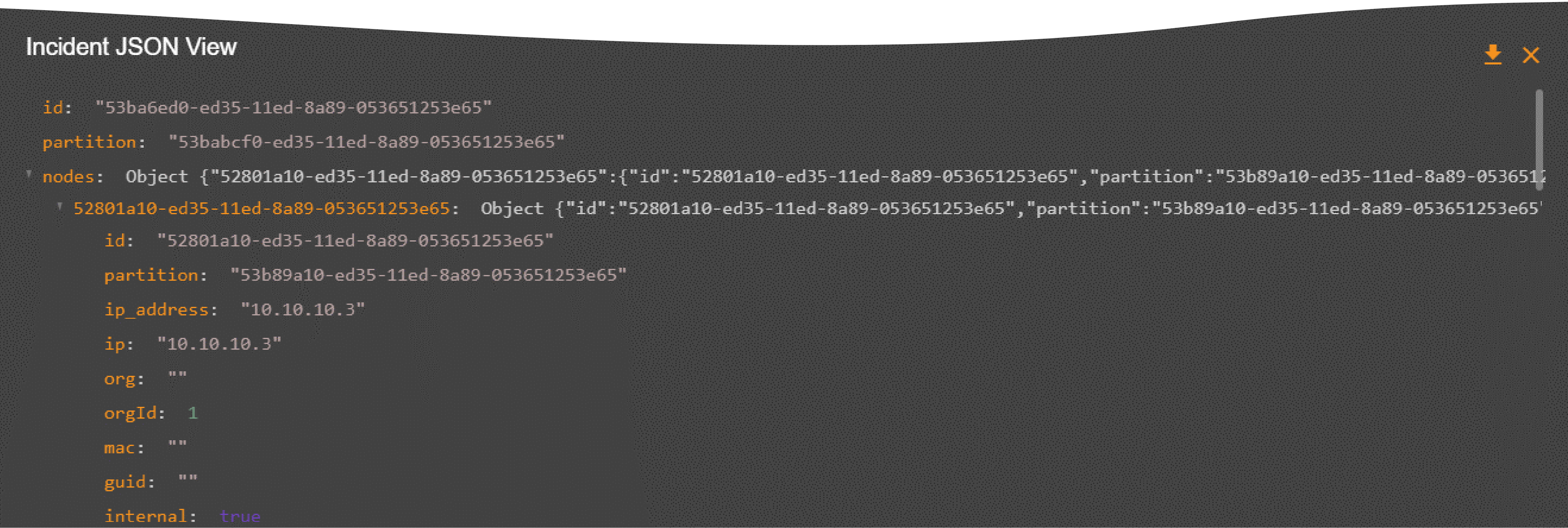
Graph vs. Crime Theory

- Meaningful Graph Relationships
- Modus Operandi of Attacker
- Combines, standardizes diverse data
- Hierarchical JSON
- *SECOPS & LE **Unit of Work***



Power of JSON

- High Compression (net & disk)
- REST Powered Transmission
- Easy to Hash & Version
- Hierarchical Structures



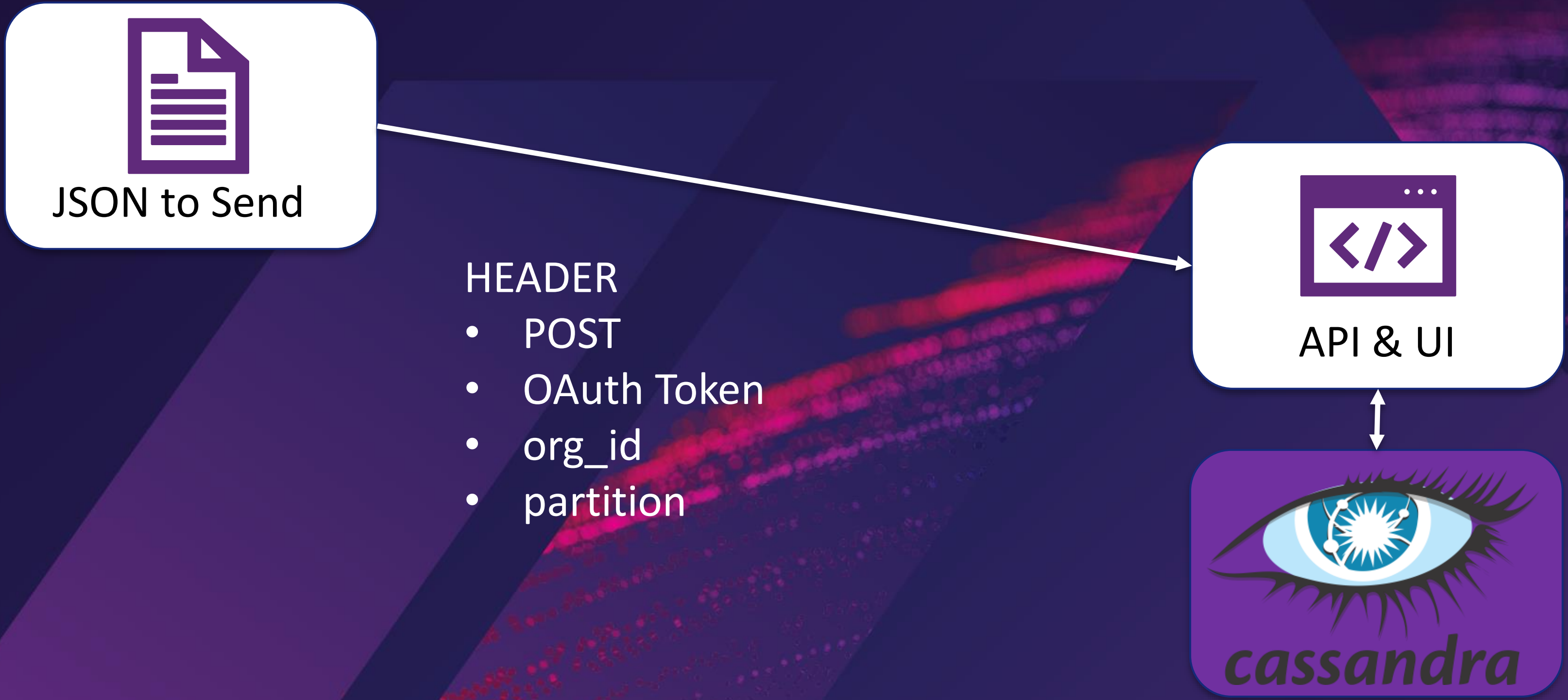
Incident JSON View

```
id:  "53ba6ed0-ed35-11ed-8a89-053651253e65"
partition:  "53babcf0-ed35-11ed-8a89-053651253e65"
nodes:  Object {"52801a10-ed35-11ed-8a89-053651253e65":{"id":"52801a10-ed35-11ed-8a89-053651253e65","partition":"53b89a10-ed35-11ed-8a89-053651
   52801a10-ed35-11ed-8a89-053651253e65:  Object {"id":"52801a10-ed35-11ed-8a89-053651253e65","partition":"53b89a10-ed35-11ed-8a89-053651253e65"
      id:  "52801a10-ed35-11ed-8a89-053651253e65"
      partition:  "53b89a10-ed35-11ed-8a89-053651253e65"
      ip_address:  "10.10.10.3"
      ip:  "10.10.10.3"
      org:  ""
      orgId:  1
      mac:  ""
      guid:  ""
      internal:  true
```

JSON Sharable Objects

- Incident Collections
- Threat Intel
- Bulletins
- Job Execution
- Reports

Standard REST Operations



JSON to Send

HEADER

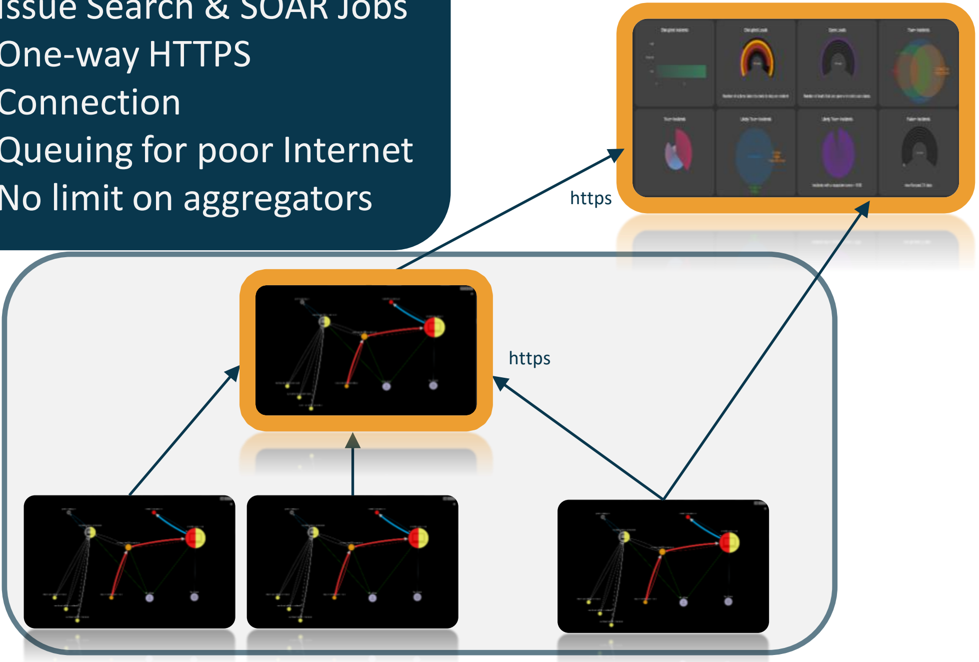
- POST
- OAuth Token
- org_id
- partition

API & UI

cassandra

WitFoo CyberGrid Federated Sharing

- Incidents, Reports, Rules
- Issue Search & SOAR Jobs
- One-way HTTPS Connection
- Queuing for poor Internet
- No limit on aggregators



Aggregate Reporting

- Cybersecurity Insurers
- Vendor Management
- Ongoing Audit
- Chain-of-Command

Aggregate Operations

- MSSP
- Military SOC
- Community SOC
- M&A
- Report to Law Enforcement

Questions & Resources

- CharlesHerring.com (talks & social)
- WitFoo.com/blog
- Community.WitFoo.com
- charles@witfoo.com

AI.dev

Open Source GenAI & ML Summit

CASSANDRA SUMMIT

#AIdotDev | #CassandraSummit

THE LINUX FOUNDATION