OpenSearchCon EUROPE

Teaching OpenSearch To Read YAML: Finding Broken Infrastructure Before It Breaks

Unnati Mishra

R & D Engg

CKA, CKAD

Public Speaker

Agenda

01 The YAML Tax

02 Infrastructure as a Graph

03 Why OpenSearch?

04 Mapping Kubernetes to OpenSearch Documents

05 Demo

06 Future Possibilities

07 Key Takeaways

"Everything was fine… until production burned."

The Incident

The Alert	The War Room	The Root Cause
3AM. PagerDuty fires. Payment service latency is spiking.	4 engineers. 2 hours. Rollbacks, restarts, flame graphs.	A YAML file. Two years old. A resource quota nobody had reviewed since the service scaled.

"We didn't have a runtime problem. We had a configuration problem that looked like a runtime problem."

frontend-service.yaml yaml . Tab: 2 spaces

```yaml
apiVersion: v1
kind: Service
metadata:
  name: frontend-service
spec:
  selector:
    app: frontend
  ports:
  - protocol: TCP
    port: 80
    targetPort: 8080
type: ClusterIP
```

Tiny Typo!

The YAML Tax

- **The Problem:** Production outages start as YAML typos.
- Errors hide in plain sight (the "Time Bomb" effect).
- We patch symptoms, not root causes.
- **The Reason:** Current tools treat configuration as "dumb text."

We debug symptoms, not causes.

Linters check Syntax. We need Intent.

- Linters (Syntax Checks):
- Catches bad indentation.
- Catches missing colons.
- Doesn't understand what the app actually does.

- Semantics (Intent Checks):
- "Does this frontend app need cluster-admin rights?"
- "Are these network policies isolating the database?"

What if YAML had meaning?

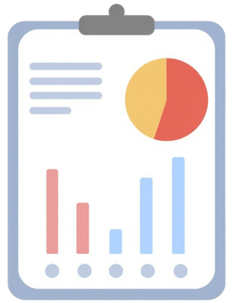
Infrastructure as a Graph

- Kubernetes is NOT a collection of flat files.
- It is a living Configuration Graph.
- Nodes: Deployments, Pods, Services, ConfigMaps.
- Edges: Selectors, Labels, RoleBindings.
- Stop counting lines. Start parsing relationships.

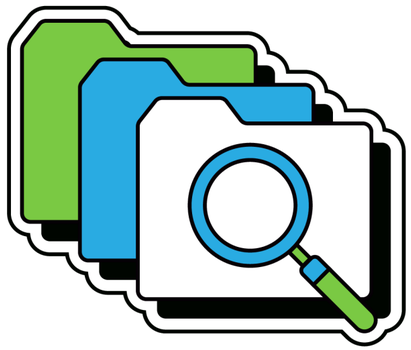
YAML → Structured Data → Searchable Graph → Insights

OpenSearch = Search Engine for Data

Why OpenSearch?

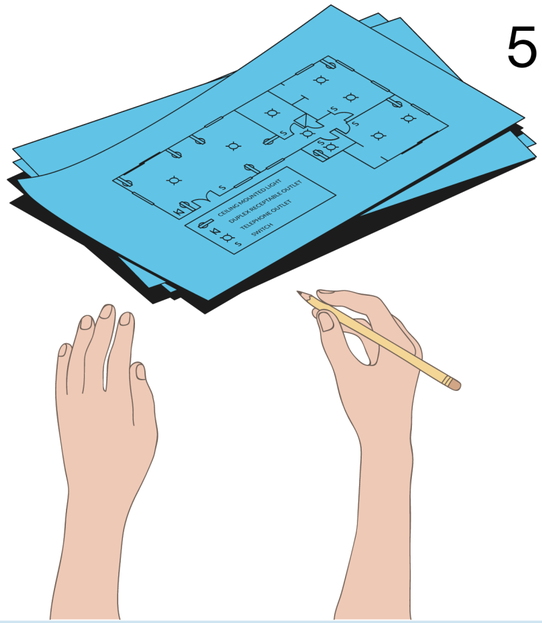


- Why a search engine for infrastructure?
- JSON Native: YAML maps perfectly to JSON documents.
- Schema-on-Write: Easily handle deeply nested Kubernetes objects.
- Aggregations: Group data logically (e.g., sum of memory limits per namespace).
- Existing Footprint: Most platform teams already run OpenSearch for logs!



The Architecture

1. **Git / CI Pipeline:** Code is committed (ArgoCD, GitHub Actions).
2. **Custom Parser:** Resolves relationships (The "Brain").
3. **JSON Conversion:** Translates enriched graph to JSON.
4. **OpenSearch Index:** Data becomes searchable in milliseconds.
5. **Policy Engine:** CI/CD queries OpenSearch to block bad PRs.



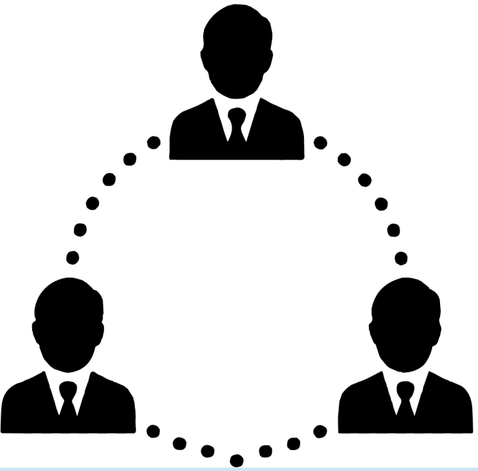
YAML → Custom Parser → JSON → OpenSearch Index → Queries

Semantic Enrichment (The Magic)

- We don't just index YAML; we enrich it.
- Raw YAML: selector: app: payment
- Enriched JSON in OpenSearch: "resolved_targets": ["payment-v1", "payment-v2"]
- Result: OpenSearch holds the answers before you even ask.

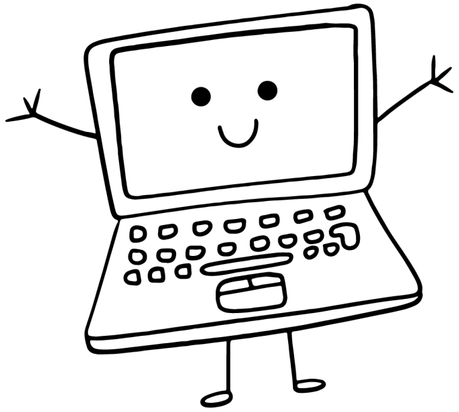
Use Case 1 - The RBAC Blast Radius

- **The Question**: "Who has cluster-admin access right now?"
- **The Old Way**: Endless kubectl commands and manual mapping.
- **The OpenSearch Way**: Query: kind: RoleBinding AND roleRef.name: cluster-admin
- **Result**: Instantly aggregate by subject. Audit complete in 50ms.



Use Case 2 - Orphaned Network Policies

- **The Question**: "Which firewall rules are pointing to deleted apps?"
- **The OpenSearch Way**: * Query: kind: NetworkPolicy AND matched_pods_count: 0
- **Result**: Instant clean-up list of dead infrastructure.
- Try doing that with a Regex.



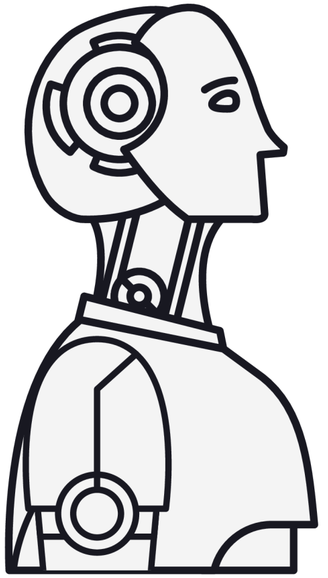
DEMO

CodesbyUnnati/OpenSearchCon-EU

The Infrastructure Immune System

- Moving from reactive patching to preventative debugging.
- OpenSearch Monitors act as automated guards:
- Flag missing memory/CPU limits.
- Flag rogue cross-namespace traffic.
- Action: Alert in Slack & Block the PR before it merges.

Future Possibilities



- **Vector Search**: Find configs structurally "similar" to known good states (kNN).
- **LLM Integration**: Feed structured OpenSearch data to AI for context-aware fixes.
- **Drift Detection**: Compare the Git OpenSearch Index vs. the Live Cluster Index.

Key Takeaways

1. Your YAML is not flat text; it's an interconnected graph.
2. Validate semantics and intent, not just syntax.
3. OpenSearch turns raw configuration into queryable infrastructure intelligence.

Stop treating YAML as text.
Start treating it as knowledge.

Thank You

Do you have any questions?

Let’s Connect!

Unnati Mishra

unnaticse2019@gmail.com

linkedin.com/in/pingunnatimishra

@ping_Unnati

Anonymous Feedback Form!


OpenSearchCon
EUROPE