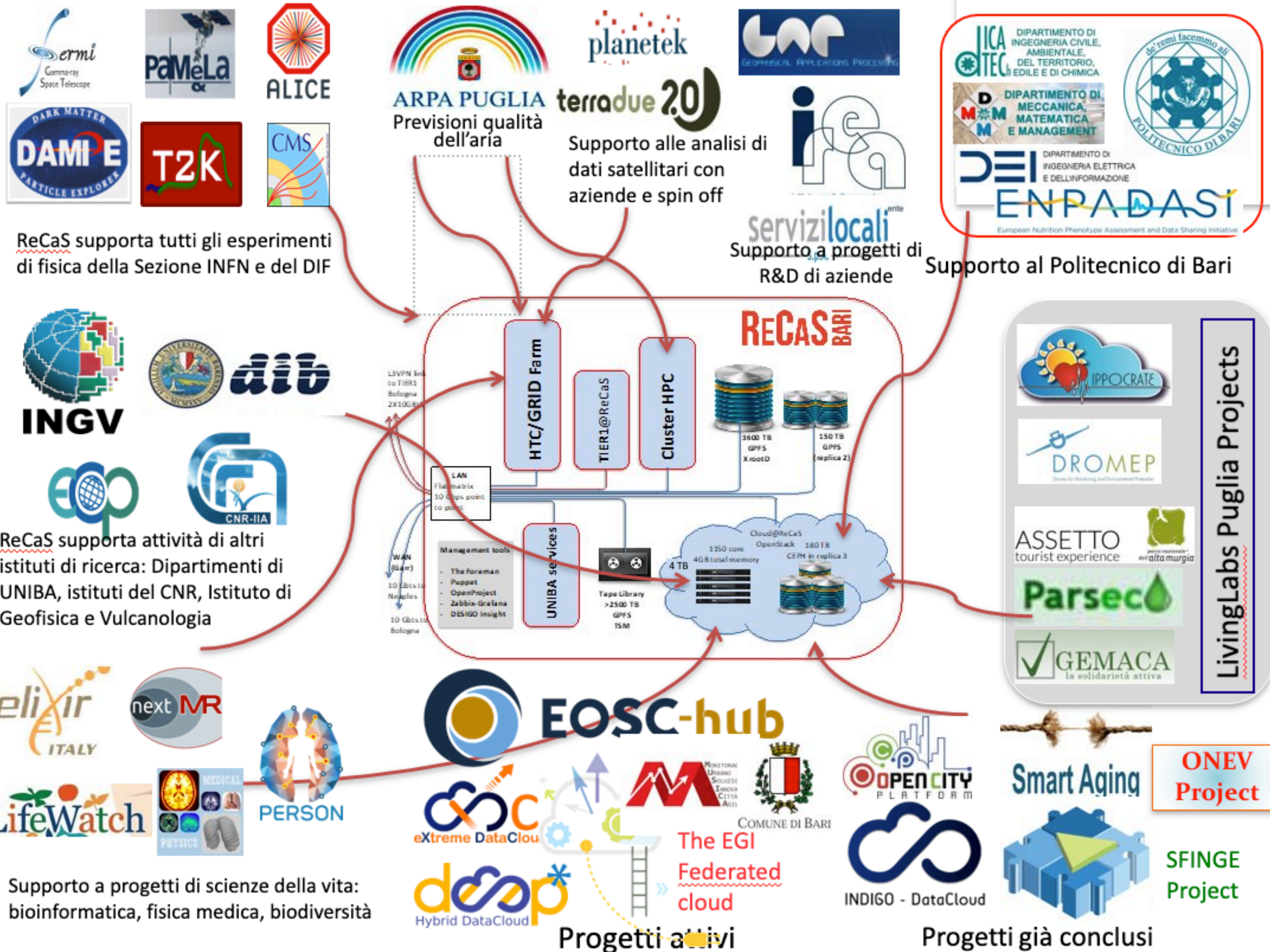


Gestione e Analisi dei Log con strumenti moderni[programmabili]

Alessandro Italiano



ReCaS supporta tutti gli esperimenti di fisica della Sezione INFN e del DIF

ReCaS supporta attività di altri istituti di ricerca: Dipartimenti di UNIBA, istituti del CNR, Istituto di Geofisica e Vulcanologia

Supporto a progetti di scienze della vita: bioinformatica, fisica medica, biodiversità

Progetti attivi

Progetti già conclusi

i log, un asset da gestire

- Il DataCenter offre ad utenti eterogenei servizi eterogenei che accedono a risorse eterogenee
 - La gestione dello stack completo include necessariamente anche quella dei log
 - l'analisi dei log e' determinate per la risoluzione delle "issues"
- I log devono essere necessariamente :
 - inviati
 - configurare servizi ed apparati per inviare i log
 - non necessariamente attraverso un unico standard/protocollo
 - collezionati
 - in modo consistente
 - analizzati
 - il concetto di analisi i log e decisamente vasto, varia dal grep all'AI

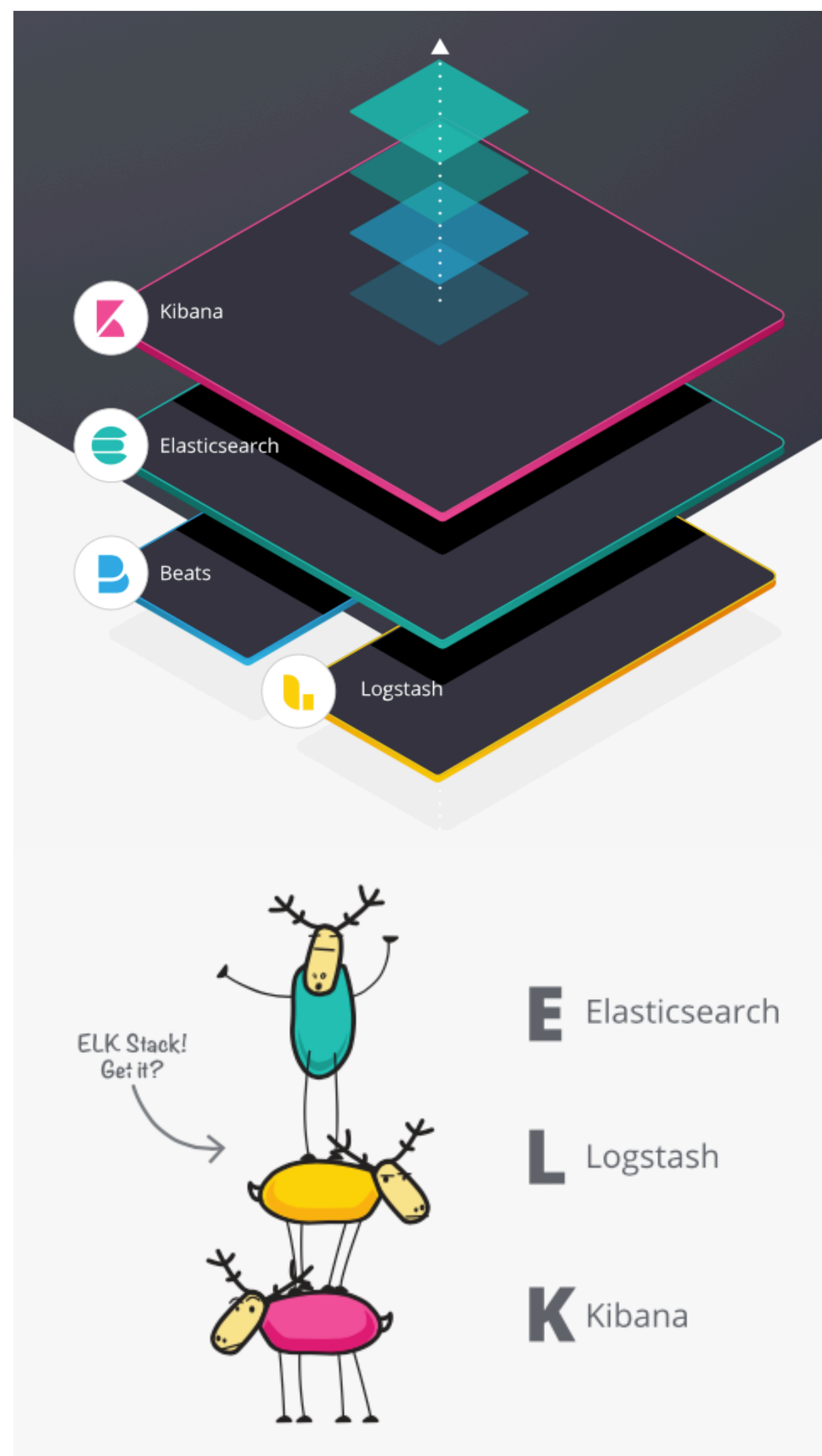
migration path

In principio era `syslog[ng]`
stabile, robusto
che scala regolarmente.
con un backend WORM

L'accesso ai log un fattore limitante

In produzione
istanza di `elasticsearch` che
processa, colleziona, indicizza
ed espone via REST API
i log del DataCenter.

elastic stack



- elasticsearch il primo componente ad essere stato rilasciato ed e' stato sviluppato "on top of"
 - Apache lucene, high-performance, full-featured search engine library written entirely in Java
 - ogni indice e' un insieme di "shards", cioe' delle singole istanze di Apache Lucene
 - in sostanza elasticsearch e' una sorta di "REST frontend server" di N istanze di "Apache lucene"
- kibana e' la dashboard dello stack che garantisce
 - accesso ai dati memorizzati
 - l'operatività ed il controllo dell'istanza di elasticsearch
- Logstash e' una "ingest pipeline"
 - input -> input processing -> output to elasticsearch
- Beats AKA[FileBeat] invia log direttamente ad elasticsearch dopo averli processati

key value

- L'elaborazione dei dati in ingresso è un valore aggiunto di elasticsearch
- attraverso tale processo, un riga di log viene trasformata in un documento formattato di tipo hash/dictionary, key:value
 - approccio standard per un accesso programmatico ai dati
 - Tale processo viene applicato lato client e quindi diventa anche architetturealmente vantaggioso
 - incide sulla qualità dell'analisi dei log
- mapping dinamico, lo schema dell'indice si adatta dinamicamente ai log che vengono inviati

Input processing

una riga di log generata su un determinato server

```
Oct 6 12:26:41 wn-1-7-22 mmfs: [N] Connecting to 90.147.169.148 wn-8-6-22.recas <con112>
```

corrisponde ad un "document" con un id univoco che non e' altro che un hash key: value

```
[root@elk-02 ~]# curl -X GET -u [REDACTED] --cacert /etc/filebeat/elasticsearch-ca.pem https://elk-02
.recas.ba.infn.it:9200/filebeat-2022.10.06-000904/_doc/g4DvrIMBWyZ9Gq_1crGW?pretty
{
  "_index" : "filebeat-2022.10.06-000904",
  "_type" : "_doc",
  "_id" : "g4DvrIMBWyZ9Gq_1crGW",
  "_version" : 1,
  "_seq_no" : 91907009,
  "_primary_term" : 1,
  "found" : true,
  "_source" : {
    "agent" : {
      "hostname" : "wn-1-7-22.recas.ba.infn.it",
      "name" : "wn-1-7-22.recas.ba.infn.it",
      "id" : "d12149af-ad57-4d8b-97e7-fdfd8176fb95",
      "type" : "filebeat",
      "ephemeral_id" : "2a970988-1ce5-4f2f-a25b-984cff6f1b55",
      "version" : "7.17.6"
    },
    "process" : {
      "name" : "mmfs"
    },
    "log" : {
      "file" : {
        "path" : "/var/log/messages"
      },
      "offset" : 1327407
    },
    "fileset" : {
      "name" : "syslog"
    },
    "message" : "[N] Connecting to 90.147.169.148 wn-8-6-22.recas <con112>",
    "tags" : [
      "wn",
      "htcondor",
      "startd",
      "syslog",
      "gpfs",
      "farm"
    ],
    "input" : {
      "type" : "log"
    },
    "@timestamp" : "2022-10-06T12:26:41.000+02:00",
    "system" : {
      "syslog" : { }
    },
    "ecs" : {
      "version" : "1.12.0"
    },
    "related" : {
      "hosts" : [
        "wn-1-7-22"
      ]
    },
    "service" : {
      "type" : "system"
    },
    "host" : {
      "hostname" : "wn-1-7-22",
      "name" : "wn-1-7-22.recas.ba.infn.it"
    },
    "event" : {
      "ingested" : "2022-10-06T10:26:53.205859199Z",
      "timezone" : "+02:00",
      "kind" : "event",
      "module" : "system",
      "dataset" : "system.syslog"
    }
  }
}
```

1. Agent Details

2. process name ricavato dinamicamente dalla riga di log

3. messaggio ricavato dinamicamente dalla riga di log

4. Tags aggiunte dall'admin

5. timestamp ricavato dinamicamente dalla riga di log

6. hostname ricavato dinamicamente dalla riga di log

7. Ingest time diverso dal timestamp

deployment

```
REcAS.com Projects Groups Activity Milestones Snippets Search or jump to...
P prebatch
Project
Repository
Files
Commits
Branches
Tags
Contributors
Graph
Compare
Charts
1 ---
2
3 recas::profiles::base::classes:
4   - elasticsearch
5   - metricbeat
6
7 recas::profiles::base::extra_pkg:
8   - jdk-12.0.2-12.0.2-ga.x86_64
9   - dstat
10
11 elasticsearch::manage_repo: false
12 elasticsearch::restart_on_change: true
13 elasticsearch::autoupgrade: true
14 elasticsearch::api_protocol: http
15 elasticsearch::api_host: "${ipaddress}"
16 elasticsearch::api_timeout: 60
17 elasticsearch::api_basic_auth_username:
18 elasticsearch::api_basic_auth_password:
19 elasticsearch::jvm_options:
20   - '-Xms32g'
21   - '-Xmx32g'
22 elasticsearch::datadir: '/elk'
23
```

1. istanzio la classe

2. definisco alcuni parametri base tra cui l'autoupgrade

3. Parametri specifici dell'istanza "logfacility"

4. il ruolo id ogni nodo può cambiare

```
24 elasticsearch::instances:
25   logfacility:
26     init_defaults:
27       MAX_OPEN_FILES: '65536'
28       MAX_LOCKED_MEMORY: 'unlimited'
29     config:
30       node.name: "${hostname}"
31       cluster.name: 'logfacility'
32       network.host: "${ipaddress}"
33       bootstrap.memory_lock: True
34       node.roles:
35         - transform
36         - data_frozen
37         - master
38         - remote_cluster_client
39         - data
40         - data_content
41         - data_warm
42         - data_cold
43         - ingest
44       discovery.zen.minimum_master_nodes: 2
45       discovery.zen.ping.unicast.hosts:
46         - "172.20.0.148"
47         - "172.20.0.10"
48         - "172.20.0.113"
49         - "172.20.0.118"
50         - "172.20.0.42"
51       xpack.security.enabled: True
52       xpack.security.audit.enabled: true
53       xpack.security.http.ssl.enabled: true
54       xpack.security.http.ssl.keystore.path: certs/elasticsearch/http.p12
55       xpack.security.transport.ssl.enabled: True
56       xpack.security.transport.ssl.verification_mode: certificate
57       xpack.security.transport.ssl.keystore.path: certs/host.p12
58       xpack.security.transport.ssl.truststore.path: certs/ca.p12
59       xpack.ml.enabled: true
60       indices.query.bool.max_clause_count: 4096
61       indices.memory.index_buffer_size: 5%
62       indices.queries.cache.size: 30%
63       thread_pool.write.size: 24
64       node.processors: 24
```

```
13   node.roles:
14     - transform
15     - data_frozen
16     - master
17     - remote_cluster_client
18     - data_content
19     - data_hot
20     - ingest
```

5. uso ssl per non inviare log in chiaro

6. ottimizzazione

- diversi metodi disponibili per fare il deployment di elasticsearch
 - puppet, ansible, package, docker, k8s
- con puppet faccio quindi il deployment di una istanza di elasticsearch servita da un cluster di macchine
- per ssl uso una self signed CA

cluster monitoring con Kibana

elastic Search Elastic Alerts and rules

Enter setup mode Clusters logfacility Last 15 minutes Show dates Refresh

Cluster overview

logfacility

Elasticsearch

Overview

- Health: Healthy
- Version: 7.17.6
- Uptime: a month
- License: Basic

Nodes: 7

- Disk Available: 29.02% (11.0 TB / 38.0 TB)
- JVM Heap: 40.63% (84.5 GB / 208.0 GB)

Indices: 534

- Documents: 74,981,774,057
- Disk Usage: 25.1 TB
- Primary Shards: 534
- Replica Shards: 480

Logs

No structured logs found. Check if the var.paths setting points to JSON logs.

Kibana

Healthy

Overview

- Requests: 0
- Max. Response Time: 588 ms

Instances: 1

- Connections: 2
- Memory Usage: 8.92% (369.7 MB / 4.0 GB)

UpTime

```
[root@elk-01 ~]# stat -c %y%z /etc/elasticsearch/logfacility/roles.yml
2019-07-23 08:47:30.752803293 +02002019-07-23 08:47:30.841802977 +0200
[root@elk-01 ~]#
```

Beats

Overview

- Total Events: 1.2k
- Bytes Sent: 1.9 MB

Beats: 7

- Metricbeat: 7

cluster monitoring con Kibana

The screenshot displays the Kibana cluster monitoring interface. At the top, the Elastic logo and navigation tabs for Clusters, logfacility, and Elasticsearch are visible. A central graph titled "elk-05.recas.ba.infn.it: Network traffic on enp3s0" shows incoming (green) and outgoing (blue) network traffic over time. Below the graph, a summary table provides statistics for both traffic types. The main dashboard area features a summary row with key metrics: Status (Green), Alerts (0), Nodes (7), Indices (534), JVM Heap (119.6 GB / 208.0 GB), Total shards (1014), Unassigned shards (0), Documents (75,033,883,029), and Data (25.1 TB). Below this is a table listing individual nodes with columns for Name, Alerts, Status, Shards, CPU Usage, Load Average, JVM Heap, and Disk Free Space. The nodes listed are elk-01 through elk-4-10-19, with elk-05 marked as a favorite.

Name ↑	Alerts	Status	Shards	CPU Usage	Load Average	JVM Heap	Disk Free Space
elk-01 172.20.0.148:9300	Clear	Online	169	↓ 1%	↑ 0.53	↑ 27%	↓ 1.0 TB
elk-02 172.20.0.113:9300	Clear	Online	169	↓ 2%	↓ 0.29	↓ 60%	↑ 2.2 TB
elk-03 172.20.0.118:9300	Clear	Online	0	↑ 15%	↑ 1.39	↓ 55%	↓ 173.8 GB
elk-04 172.20.0.10:9300	Clear	Online	169	↑ 6%	↓ 1.33	↓ 45%	↑ 2.9 TB
★ elk-05 172.20.0.42:9300	Clear	Online	169	↓ 7%	↓ 1.38	↑ 65%	↑ 2.3 TB
elk-1-1-19 172.20.18.33:9300	Clear	Online	169	↑ 13%	↓ 2.27	↓ 45%	↑ 1.2 TB
elk-4-10-19 172.20.168.172:9300	Clear	Online	169	↓ 14%	↓ 1.97	↓ 46%	↑ 1.2 TB

send data using FileBeat

- il modo piu' naturale per inviare logs ad elasticsearch e' attraverso FileBeat
- attraverso puppet configuro FileBeat per inviare i dati al cluster
- FileBeat viene distribuito con dei moduli che possono essere richiamati nella configurazione per inviare classi di log note.
- il modulo quindi gestisce l' "input processing" dei log più comuni in modo che ad ogni riga di log corrisponda una entry nell'indice, cioè un dizionario key: value
- in caso di problemi nell'invio dei log FileBeat ricomincia dall'ultimo inviato
- posso inviare anche log presenti in custom files [1] [2]

[1]

```
249 filebeat::inputs:
250   webdav-access:
251     paths:
252       - "/var/log/storm/webdav/storm-webdav-*.log"
253     fields:
254       storm: true
255       webdav: true
256     fields_under_root: true
257     tail_files: true
258
259
```

[2]

```
47 filebeat::inputs:
48   condor:
49     paths:
50       - "/var/log/condor/*Log"
51     fields:
52       htcondor: true
53       schedd: true
54     fields_under_root: true
55     tail_files: true
56   cream:
57     paths:
58       - "/var/log/condor-ce/*Log"
59     fields:
60       condorce: true
61       htcondor: true
62     fields_under_root: true
63     tail_files: true
64
```

```
41 filebeat::service_ensure: running
42 filebeat::package_ensure: latest
43
44 filebeat::outputs:
45   elasticsearch:
46     hosts:
47       - "https://elk-02.recas.ba.infn.it:9200"
48       - "https://elk-03.recas.ba.infn.it:9200"
49       - "https://elk-04.recas.ba.infn.it:9200"
50       - "https://elk-05.recas.ba.infn.it:9200"
51     ssl_certificate_authorities: ["/etc/filebeat/elasticsearch-ca.pem"]
52     loadbalance: True
53     api_key: "StvzYXYBUG_eyB7hb96K:x56vYfWHRDWkq88BRi7-LQ"
54     bulk_max_size: 50
55
56 filebeat::setup:
57   ilm:
58     enabled: "auto"
59     rollover_alias: "filebeat"
60     pattern: "{now/d}-000001"
61     policy_name: "filebeat-7.2.0"
62   template:
63     name: "filebeat"
64     pattern: "filebeat-*"
65     settings:
66       index.number_of_replicas: 1
67
68 filebeat::modules:
69   - module: auditd
70     log:
71       input:
72         tail_files: true
73   - module: system
74     syslog:
75       input:
76         tail_files: true
77 filebeat::logging:
78   to_files: true
79   files:
80     path: /var/log/filebeat
81     name: filebeat
82     keepfiles: 1
83     permissions: 0644
84
```

1. Distribuisco il carico su più nodi

2. Connessione sicura

3. uso "api_key" definita per lo scopo evitando di distribuire username/pwd

4. posso fare anche il setup dell'indice su cui fare l'upload dei log

5. uso due moduli di filebeat per due classi di log note, syslog ed audit

6. i log di filebeat li metto su un file diverso da syslog

send data from custom source

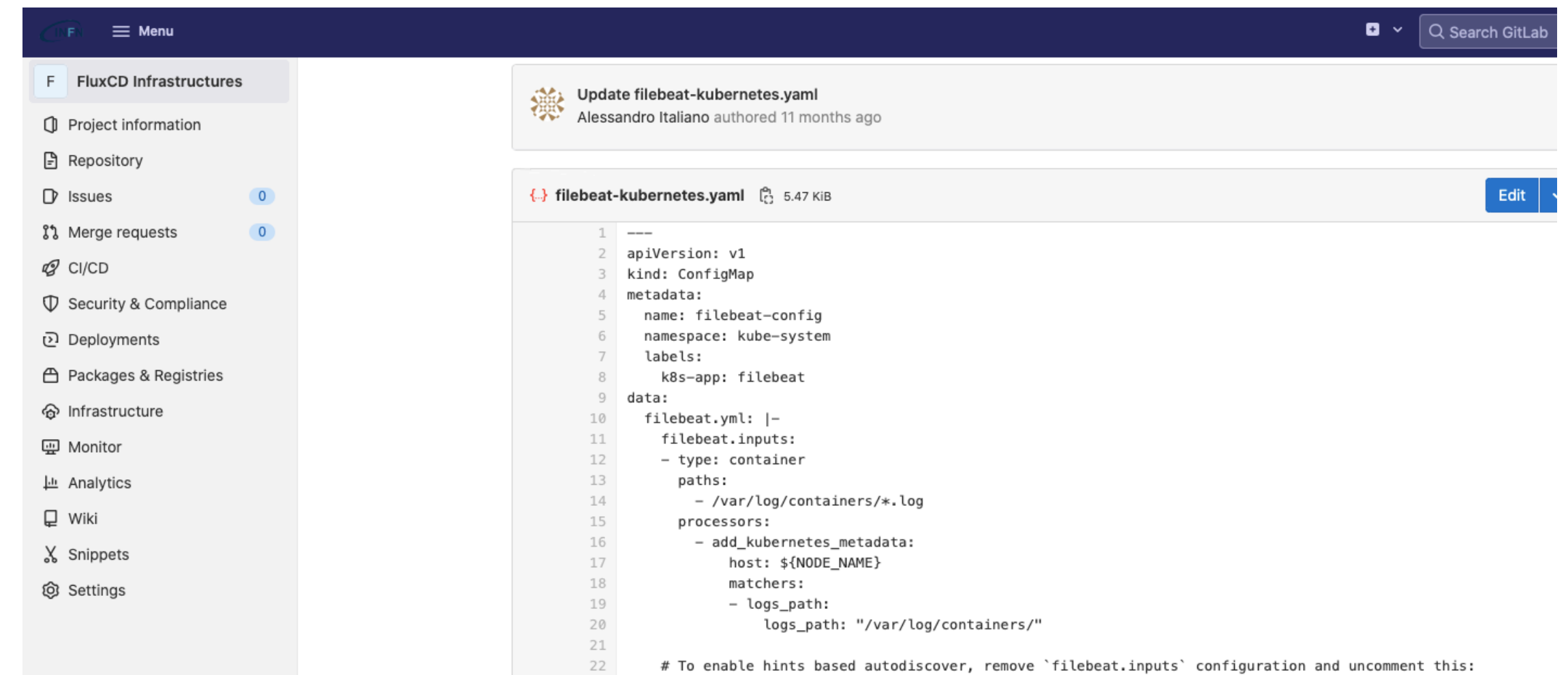
Sfruttando Logstash riesco ad inviare i log anche dai dispositivi che non consentono l'installazione di FileBeat, nel caso specifico Logstash :

- 1.si mette in ascolto sulla porta 162/UDP in ascolto per eventuali SNMP Trap
 - 2.si mette in ascolto sulla porta 514/UDP in ascolto per eventuali messaggi inviati da syslog
- in entrambi in casi Logstash processai dati ricevuti formattandoli ed inviandoli all'indice

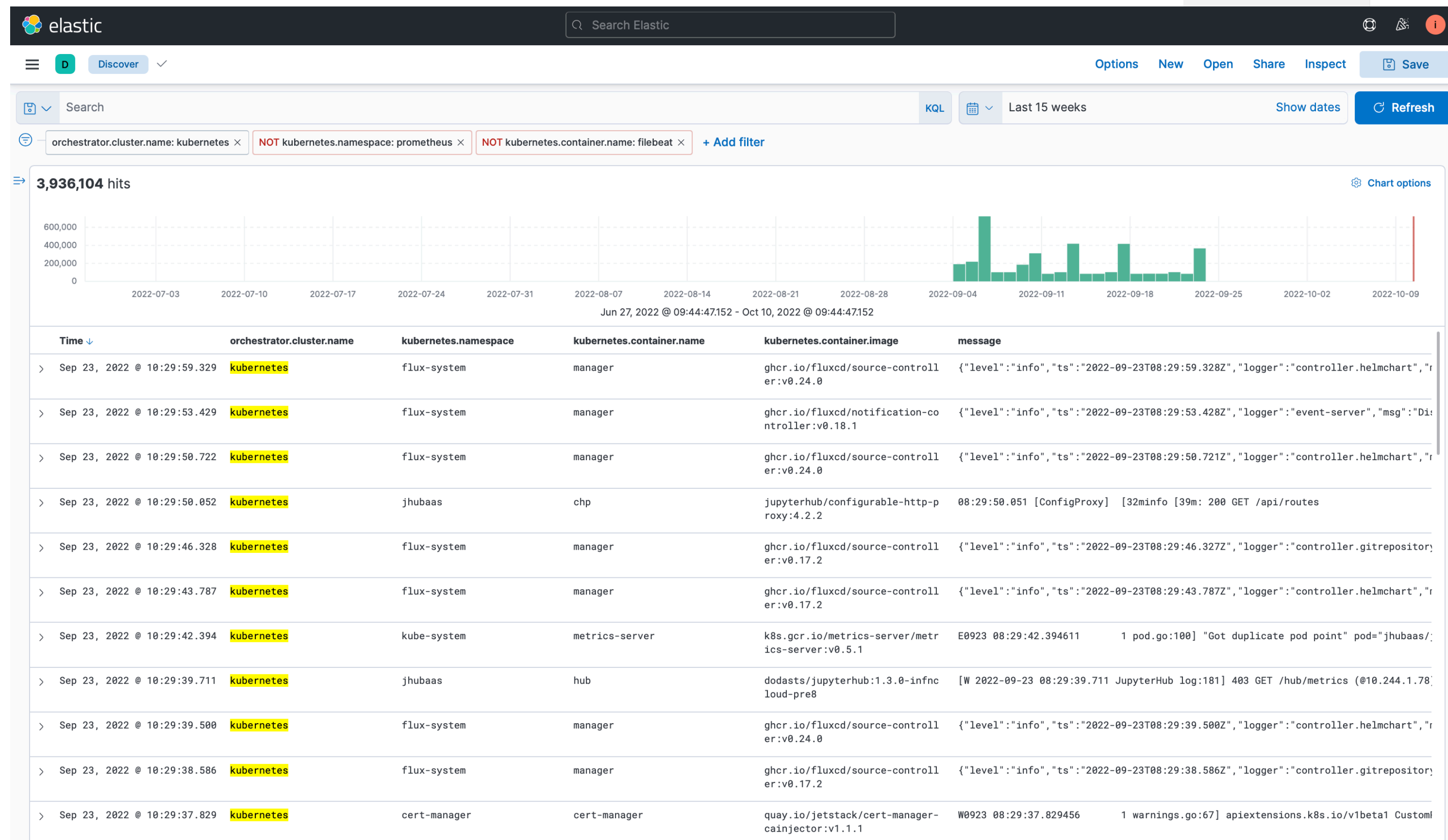
```
14 if $facts['fqdn'] == 'elk-03.recas.ba.infn.it' {
15   logstash::configfile { 'snmptrap':
16     content => 'input { snmptrap
17                   { type => snmptrap
18                     host => "172.20.0.118"
19                     port => "162"
20                     codec => "json"
21                     community => ["LXCA_COMMUNITY", "storage_public", "storage_private", "dell_storage", "ce12800",
22                   output { elasticsearch { hosts => ["https://172.20.0.118:9200"]
23                                     cacert => "/etc/logstash/elasticsearch-ca.pem"
24                                     user => "elastic" password => "[REDACTED]" } }',
25   }
26   logstash::configfile { 'syslog':
27     content => ' input { udp
28                   { port => 514
29                     type => "syslog" }
30   }
31   output { elasticsearch { hosts => ["https://172.20.0.118:9200"]
32                                     cacert => "/etc/logstash/elasticsearch-ca.pem"
33                                     user => "elastic" password => "[REDACTED]" } }',
34   }
35 }
36 }
```

send data from k8s

- su k8s posso lanciare un container/app che fa partire filebeat
- invio dei log file di tutti i container/app in esecuzione su k8s



The screenshot shows a GitLab repository interface for a project named "FluxCD Infrastructures". The left sidebar contains a navigation menu with options like Project information, Repository, Issues, Merge requests, CI/CD, Security & Compliance, Deployments, Packages & Registries, Infrastructure, Monitor, Analytics, Wiki, Snippets, and Settings. The main content area displays the configuration file "filebeat-kubernetes.yaml" with a file size of 5.47 KiB. The configuration is a ConfigMap in the kube-system namespace, defining filebeat inputs and processors for container logs. It includes a section for Elasticsearch output and a DaemonSet spec for deployment.



The screenshot shows the Elastic search interface. The search bar contains the query: `orchestrator.cluster.name: kubernetes NOT kubernetes.namespace: prometheus NOT kubernetes.container.name: filebeat`. The results show 3,936,104 hits. A bar chart displays the distribution of hits over time, with a peak around September 4, 2022. Below the chart is a table of search results with columns for Time, orchestrator.cluster.name, kubernetes.namespace, kubernetes.container.name, kubernetes.container.image, and message.

Time ↓	orchestrator.cluster.name	kubernetes.namespace	kubernetes.container.name	kubernetes.container.image	message
Sep 23, 2022 @ 10:29:59.329	kubernetes	flux-system	manager	ghcr.io/fluxcd/source-controller:v0.24.0	{"level": "info", "ts": "2022-09-23T08:29:59.328Z", "logger": "controller.helmchart", "r
Sep 23, 2022 @ 10:29:53.429	kubernetes	flux-system	manager	ghcr.io/fluxcd/notification-controller:v0.18.1	{"level": "info", "ts": "2022-09-23T08:29:53.428Z", "logger": "event-server", "msg": "Di
Sep 23, 2022 @ 10:29:50.722	kubernetes	flux-system	manager	ghcr.io/fluxcd/source-controller:v0.24.0	{"level": "info", "ts": "2022-09-23T08:29:50.721Z", "logger": "controller.helmchart", "r
Sep 23, 2022 @ 10:29:50.052	kubernetes	jhubaas	chp	jupyterhub/configurable-http-proxy:4.2.2	08:29:50.051 [ConfigProxy] [32minfo [39m: 200 GET /api/routes
Sep 23, 2022 @ 10:29:46.328	kubernetes	flux-system	manager	ghcr.io/fluxcd/source-controller:v0.17.2	{"level": "info", "ts": "2022-09-23T08:29:46.327Z", "logger": "controller.gitrepository,
Sep 23, 2022 @ 10:29:43.787	kubernetes	flux-system	manager	ghcr.io/fluxcd/source-controller:v0.17.2	{"level": "info", "ts": "2022-09-23T08:29:43.787Z", "logger": "controller.helmchart", "r
Sep 23, 2022 @ 10:29:42.394	kubernetes	kube-system	metrics-server	k8s.gcr.io/metrics-server/metrics-server:v0.5.1	E0923 08:29:42.394611 1 pod.go:100] "Got duplicate pod point" pod="jhubaas/
Sep 23, 2022 @ 10:29:39.711	kubernetes	jhubaas	hub	dodasts/jupyterhub:1.3.0-infncloud-pre8	[W 2022-09-23 08:29:39.711 JupyterHub log:181] 403 GET /hub/metrics (@10.244.1.78:
Sep 23, 2022 @ 10:29:39.500	kubernetes	flux-system	manager	ghcr.io/fluxcd/source-controller:v0.24.0	{"level": "info", "ts": "2022-09-23T08:29:39.500Z", "logger": "controller.helmchart", "r
Sep 23, 2022 @ 10:29:38.586	kubernetes	flux-system	manager	ghcr.io/fluxcd/source-controller:v0.17.2	{"level": "info", "ts": "2022-09-23T08:29:38.586Z", "logger": "controller.gitrepository,
Sep 23, 2022 @ 10:29:37.829	kubernetes	cert-manager	cert-manager	quay.io/jetstack/cert-manager-cainjector:v1.1.1	W0923 08:29:37.829456 1 warnings.go:67] apiextensions.k8s.io/v1beta1 Customl



The screenshot shows the configuration file "filebeat-kubernetes.yaml" in a code editor. The configuration is a ConfigMap in the kube-system namespace, defining filebeat inputs and processors for container logs. It includes a section for Elasticsearch output and a DaemonSet spec for deployment.

```
---
2 apiVersion: v1
3 kind: ConfigMap
4 metadata:
5   name: filebeat-config
6   namespace: kube-system
7   labels:
8     k8s-app: filebeat
9 data:
10  filebeat.yml: |-
11    filebeat.inputs:
12    - type: container
13      paths:
14        - /var/log/containers/*.log
15    processors:
16      - add_kubernetes_metadata:
17          host: ${NODE_NAME}
18          matchers:
19            - logs_path:
20                logs_path: "/var/log/containers/"
21
22  # To enable hints based autodiscover, remove `filebeat.inputs` configuration and uncomment this:
23  #filebeat.autodiscover:
24  # providers:
25  #   - type: kubernetes
26  #     node: ${NODE_NAME}
27  #     hints.enabled: true
28  #     hints.default_config:
29  #       type: container
30  #       paths:
31  #         - /var/log/containers/*${data.kubernetes.container.id}.log
32
33  processors:
34    - add_cloud_metadata:
35    - add_host_metadata:
36
37  output.elasticsearch:
38    hosts: ["https://${ELASTICSEARCH_HOST}:elasticsearch:${ELASTICSEARCH_PORT:9200}"]
39    api_key: "N2t3HXwBrezoymovrinw:HiJ-rI8KSgajLgKjBvCYlQ"
40    index: "infnccloud-logs"
41  setup:
42    ilm:
43      enabled: false
44      rollover_alias: "infnccloud-logs"
45      pattern: "{now/d}-000001"
46  template:
47    enabled: true
48    name: "filebeat"
49    pattern: "infnccloud-logs-*"
50    settings:
51      index.number_of_replicas: 0
52
53 ---
54
55 apiVersion: apps/v1
56 kind: DaemonSet
57 metadata:
58   name: filebeat
59   namespace: kube-system
60   labels:
61     k8s-app: filebeat
62 spec:
63   selector:
64     matchLabels:
65       k8s-app: filebeat
66   template:
67     metadata:
68       labels:
69         k8s-app: filebeat
70   spec:
71     serviceAccountName: filebeat
72     terminationGracePeriodSeconds: 30
73     hostNetwork: true
74     dnsPolicy: ClusterFirstWithHostNet
75     containers:
76     - name: filebeat
77       image: docker.elastic.co/beats/filebeat:7.15.0
78       args: [
79         "-c", "/etc/filebeat.yml",
80         "-e",
81       ]
82     env:
```

send data from zeek

- ZEEK e' in Intrusion Detection System [IDS]
- logga le informazione relative alle connessione che vede sulla porta dove fa sniff
- dopo aver configurato il formato dei log di tipo JSON
- attraverso un modulo specifico di filebeat invio i dati su un indice dedicato

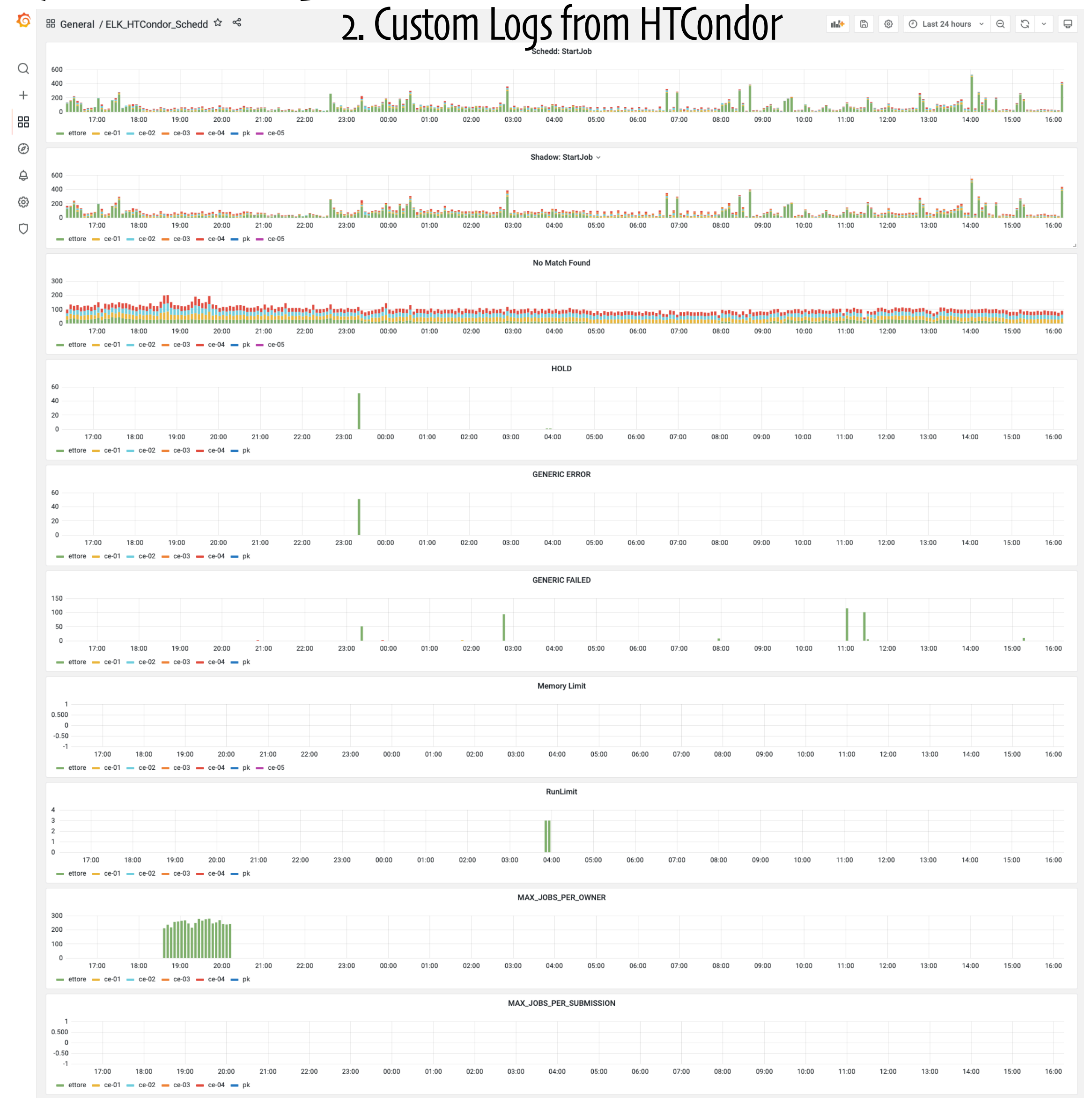
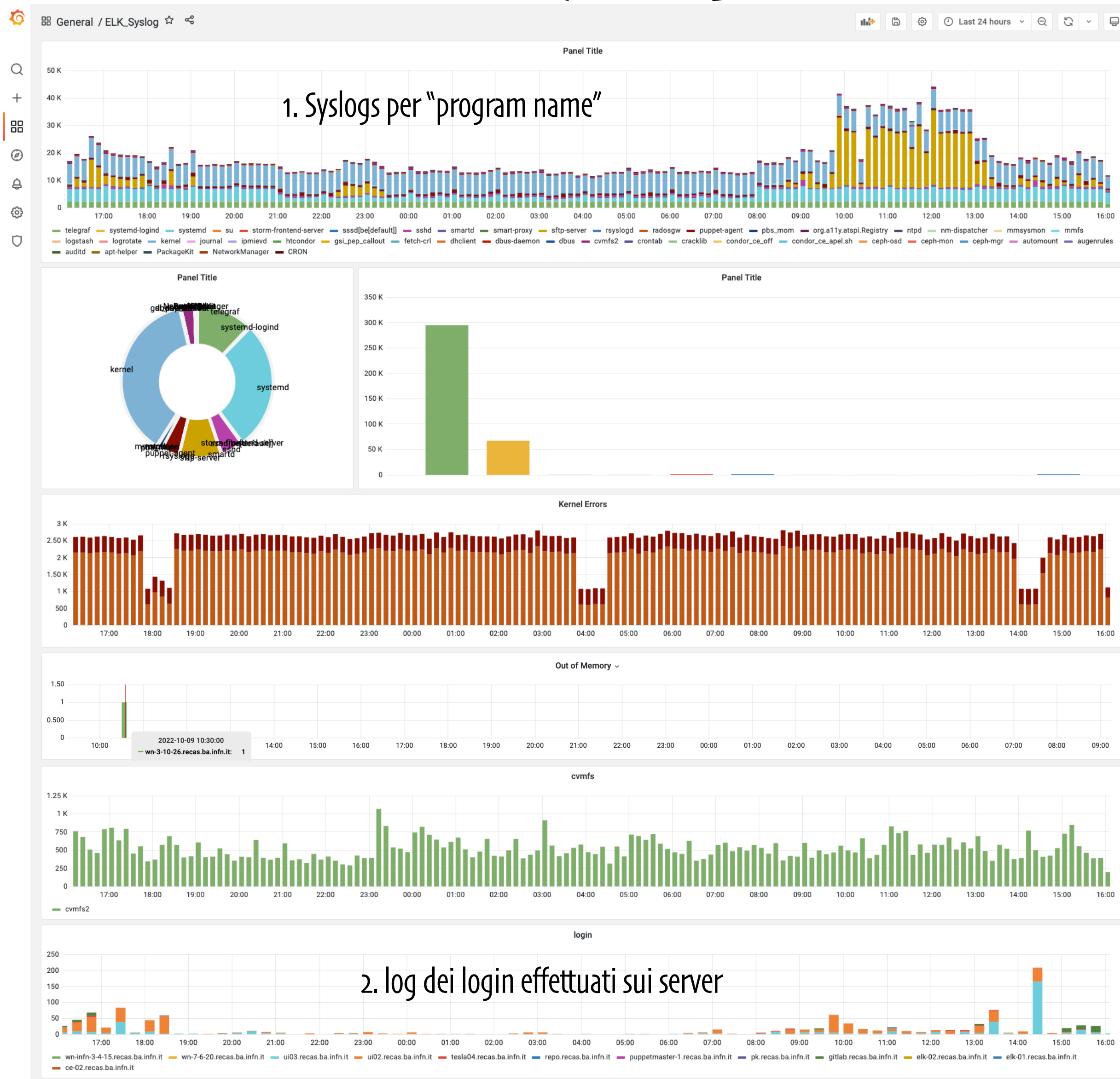
```
15 filebeat::setup:
16   ilm:
17     enabled: "auto"
18     rollover_alias: "zeek"
19     pattern: "{now/d}-000001"
20   template:
21     name: "zeek"
22     pattern: "zeek-*"
23
24
25 filebeat::modules:
26 - module: zeek
27   notice:
28     var.paths:
29       - "/usr/local/zeek/logs/current/notice.log"
30       - "/usr/local/zeek/logs/current/weird.log"
31     input:
32       tail_files: true
33   connection:
34     var.paths:
35       - "/usr/local/zeek/logs/current/conn.log"
36     input:
37       tail_files: true
38   ssl:
39     var.paths:
40       - "/usr/local/zeek/logs/current/ssl.log"
41     input:
42       tail_files: true
43   http:
44     var.paths:
45       - "/usr/local/zeek/logs/current/http.log"
46     input:
47       tail_files: true
48 - module: auditd
49   log:
50     input:
51       tail_files: true
52 - module: system
53   syslog:
54     input:
55       tail_files: true
56
```

Index Lifecycle Management [ILM] Policy

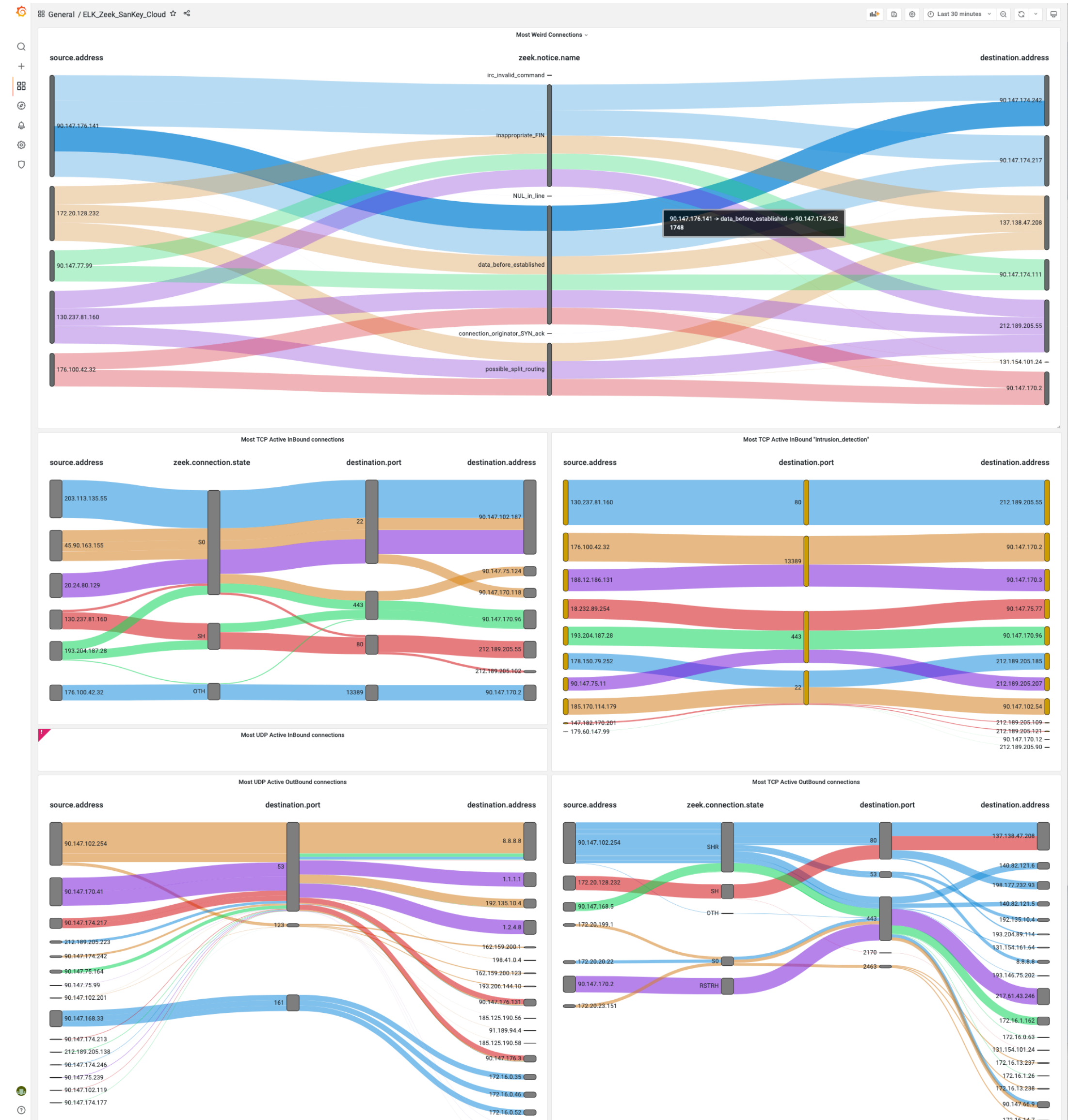
The screenshot shows the Elastic Stack Management console for editing an ILM policy named 'filebeat-7.2.0'. The interface includes a left-hand navigation menu with categories like Management, Data, Alerts and Insights, Security, Kibana, and Stack. The main content area is titled 'Edit policy filebeat-7.2.0' and contains a 'Policy summary' section. This summary shows a horizontal bar representing the policy's phases: Hot phase (red), Warm phase (yellow), and Delete phase (grey). Below this, the 'Hot phase' is expanded, showing its description and a link to 'Advanced settings'. The 'Warm phase' is also expanded, showing its description, a toggle switch, and configuration options for moving data into the phase (2 days old). The 'Delete phase' is expanded, showing its description and configuration options (180 days old). At the bottom, there is a 'Wait for snapshot policy' section with a text input field and a 'Save policy' button.

- per ogni indice possiamo definire una policy per la gestione automatica dei dati
- retention, teniamo i dati per N giorni e poi li cancelliamo in modo da essere GDPR compliance
- Fase HOT e WARM per tenere i dati piu' recenti su macchine con risorse piu' performanti
- rollover per creare un indice ogni giorno/size
 - aumenta la probabilità di avere dati piu' recenti su un indice di piccole dimensioni
 - un indice alias per accedere a tutti i dati in maniera trasparente

syslog data analysis con grafana



zeek data analysis con grafana



data analysis con kibana

The screenshot shows the Kibana search interface. The search bar contains the query `host: 172.16.0.219` and `fortigate_subtype: ips`. The results are displayed as a table with 1,374 hits. A bar chart above the table shows the distribution of hits over time, with a peak around October 5, 2022. The table columns include `Time`, `fortigate_subtype`, `attack`, `action`, `crlevel`, `crscore`, `geosrcip.ip`, and `geosrcip.timezone`.

Time	fortigate_subtype	attack	action	crlevel	crscore	geosrcip.ip	geosrcip.timezone
Oct 6, 2022 @ 16:45:36.000	ips	UPnP.SSDP.M.Search.Anomaly	dropped	low	5	85.114.131.200	Europe/Berlin
Oct 6, 2022 @ 16:45:36.000	ips	UPnP.SSDP.M.Search.Anomaly	dropped	low	5	85.114.131.200	Europe/Berlin
Oct 6, 2022 @ 16:32:36.000	ips	SSH.Client.Request.Mimicking	dropped	low	5	116.193.159.2	Asia/Hong_Kong
Oct 6, 2022 @ 16:32:36.000	ips	SSH.Client.Request.Mimicking	dropped	low	5	116.193.159.2	Asia/Hong_Kong
Oct 6, 2022 @ 16:30:58.000	ips	SNMP.Spec.Violation	dropped	low	5	157.245.35.108	Europe/London
Oct 6, 2022 @ 16:30:58.000	ips	SNMP.Spec.Violation	dropped	low	5	157.245.35.108	Europe/London
Oct 6, 2022 @ 16:30:16.000	ips	UPnP.SSDP.M.Search.Anomaly	dropped	low	5	165.22.114.105	Europe/London
Oct 6, 2022 @ 16:30:16.000	ips	UPnP.SSDP.M.Search.Anomaly	dropped	low	5	165.22.114.105	Europe/London
Oct 6, 2022 @ 15:03:24.000	ips	ZGrab.Scanner	dropped	low	5	139.59.66.9	Asia/Kolkata
Oct 6, 2022 @ 15:03:24.000	ips	ZGrab.Scanner	dropped	low	5	139.59.66.9	Asia/Kolkata
Oct 6, 2022 @ 15:03:14.000	ips	HTTP.Unknown.Tunnelling	dropped	-	-	139.59.66.9	Asia/Kolkata
Oct 6, 2022 @ 15:03:14.000	ips	HTTP.Unknown.Tunnelling	dropped	-	-	139.59.66.9	Asia/Kolkata
Oct 6, 2022 @ 14:36:20.000	ips	DNS.Invalid.OPcode	dropped	-	-	37.44.238.145	Europe/Paris
Oct 6, 2022 @ 14:36:20.000	ips	DNS.Invalid.OPcode	dropped	-	-	37.44.238.145	Europe/Paris
Oct 6, 2022 @ 14:11:38.000	ips	Nmap.Script.Scanner	dropped	low	5	43.131.68.225	Europe/Moscow
Oct 6, 2022 @ 14:11:38.000	ips	Nmap.Script.Scanner	dropped	low	5	43.131.68.225	Europe/Moscow
Oct 6, 2022 @ 14:11:30.000	ips	Nmap.Script.Scanner	dropped	low	5	43.131.68.225	Europe/Moscow
Oct 6, 2022 @ 14:11:30.000	ips	Nmap.Script.Scanner	dropped	low	5	43.131.68.225	Europe/Moscow
Oct 6, 2022 @ 13:38:09.000	ips	SSH.Client.Request.Mimicking	dropped	low	5	20.171.106.5	America/Phoenix
Oct 6, 2022 @ 13:38:09.000	ips	SSH.Client.Request.Mimicking	dropped	low	5	20.171.106.5	America/Phoenix
Oct 6, 2022 @ 13:06:54.000	ips	Linux.Kernel.TCP.SACK.Panic.DoS	dropped	high	30	131.159.24.205	Europe/Berlin
Oct 6, 2022 @ 13:06:54.000	ips	Linux.Kernel.TCP.SACK.Panic.DoS	dropped	high	30	131.159.24.205	Europe/Berlin
Oct 6, 2022 @ 12:41:09.000	ips	Censys.io.Scanner	dropped	low	5	167.248.133.120	America/Los_Angeles
Oct 6, 2022 @ 12:41:09.000	ips	Censys.io.Scanner	dropped	low	5	167.248.133.120	America/Los_Angeles
Oct 6, 2022 @ 12:13:18.000	ips	Censys.io.Scanner	dropped	low	5	167.94.138.120	America/Chicago
Oct 6, 2022 @ 12:13:18.000	ips	Censys.io.Scanner	dropped	low	5	167.94.138.120	America/Chicago
Oct 6, 2022 @ 10:41:22.000	ips	Nmap.Script.Scanner	dropped	low	5	173.230.144.72	America/Los_Angeles
Oct 6, 2022 @ 10:41:22.000	ips	Nmap.Script.Scanner	dropped	low	5	173.230.144.72	America/Los_Angeles
Oct 6, 2022 @ 10:13:23.000	ips	UPnP.SSDP.M.Search.Anomaly	dropped	low	5	64.62.197.180	America/Los_Angeles
Oct 6, 2022 @ 10:13:23.000	ips	UPnP.SSDP.M.Search.Anomaly	dropped	low	5	64.62.197.180	America/Los_Angeles
Oct 6, 2022 @ 10:07:36.000	ips	D-Link.Devices.HNAP.SOAPAction-Header.Command.Execution	dropped	critical	50	178.72.69.216	Asia/Yekaterinburg
Oct 6, 2022 @ 10:07:36.000	ips	D-Link.Devices.HNAP.SOAPAction-Header.Command.Execution	dropped	critical	50	178.72.69.216	Asia/Yekaterinburg

The screenshot shows the Kibana Inspector interface. The search bar contains the query `agent.hostname: wn-8-9-8.recas.ba.inf.it`. The results are displayed as a table with 711,390 hits. A bar chart above the table shows the distribution of hits over time, with a peak around October 4, 2022. The table columns include `Time` and `message`.

Time	message
Oct 5, 2022 @ 01:19:33.000	megaraid_sas 0000:01:00:0: Init cmd return status FAILED for SCSI host 0
Oct 5, 2022 @ 01:18:49.313	10/05/22 01:18:43 About to update statistics in shared_port daemon ad file at /var/lock/condor/shared_port_ad :
Oct 5, 2022 @ 01:18:49.313	ForkedChildrenCurrent = 0
Oct 5, 2022 @ 01:18:49.313	ForkedChildrenPeak = 0
Oct 5, 2022 @ 01:18:49.313	MyAddress = "<90.147.168.230:9618?addr=90.147.168.230-9618&alias=wn-8-9-8.recas.ba.inf.it&noUDP>"
Oct 5, 2022 @ 01:18:49.313	RequestsBlocked = 593
Oct 5, 2022 @ 01:18:49.313	RequestsFailed = 593
Oct 5, 2022 @ 01:18:49.313	RequestsPendingCurrent = 2
Oct 5, 2022 @ 01:18:49.313	RequestsPendingPeak = 16
Oct 5, 2022 @ 01:18:49.313	RequestsSucceeded = 392370
Oct 5, 2022 @ 01:18:49.313	SharedPortCommandSInfuls = "<90.147.168.230:9618?alias=wn-8-9-8.recas.ba.inf.it>"
Oct 5, 2022 @ 01:16:24.000	megaraid_sas 0000:01:00:0: Current firmware supports maximum commands: 928#011 LDIO threshold: 237
Oct 5, 2022 @ 01:16:24.000	megaraid_sas 0000:01:00:0: Performance mode :Latency
Oct 5, 2022 @ 01:16:24.000	megaraid_sas 0000:01:00:0: megasas_disable_intr_fusion is called outbound_intr_mask:0x40000009
Oct 5, 2022 @ 01:16:24.000	megaraid_sas 0000:01:00:0: FW now in Ready state
Oct 5, 2022 @ 01:16:24.000	megaraid_sas 0000:01:00:0: FW now in Ready state
Oct 5, 2022 @ 01:16:24.000	megaraid_sas 0000:01:00:0: FW supports sync cache#011: No
Oct 5, 2022 @ 01:16:16.000	megaraid_sas 0000:01:00:0: Waiting for FW to come to ready state
Oct 5, 2022 @ 01:16:10.000	megaraid_sas 0000:01:00:0: Init cmd return status FAILED for SCSI host 0

data analysis con Anomaly Detection

- Attraverso un algoritmo di Anomaly Detection chiamato IsolationForest volevo analizzare i dati relativi alle connessioni loggate da zeek per mettere in evidenza connessioni anomale dal punto di vista dei FLAGS TCP, usando python

```
53 context = create_default_context(cafile="/etc/filebeat/elasticsearch-ca.pem")
54 es = Elasticsearch(['elk-01.recas.ba.infn.it'],
55                 http_auth=
56                 scheme="https",
57                 port=9200,
58                 ssl_context=context,)
59
```

1. definisco la connessione all'istanza di elk

```
63 body = { "_source":["@timestamp",
64                 "source.address",
65                 "destination.port",
66                 'destination.ip',
67                 'zeek.connection.history',
68                 'zeek.connection.state',
69                 "source.packets",
70                 "source.bytes"],
71
```

2. scelgo le keys da usare per l'analisi

```
72 "size": 10000, "sort": [{"@timestamp": {"order": "desc"}},
73 "query":
74 {"bool":
75 {"must":
76 [{"match": {"fileset.name": "connection"}},
77 {"match": {"destination.ip": "90.147.170.3"}},
78 {"match": {"network.direction": "inbound"}},
79 {"match": {"network.transport": "tcp"}},
80 {"range": {"destination.bytes": {"gte": 1}}}
81 ],
82 "filter": [{"range": {"@timestamp": {"gte": "now-5m"}}}]}
83 }
```

3. restringo i risultati della query

4. eseguo la ricerca

```
88 response = es.search(index='zeek', scroll='1m', body=body, request_timeout=600)
107 data = es.scroll(scroll_id=sid, scroll='2m', request_timeout=600)
116 es_docs = es_docs + data['hits']['hits']
130 es_df = pandas.io.json.json_normalize(es_docs).dropna()
```

5. creo un Panda dataframe coi risultati della ricerca



esempio completo lo trovate qui: <https://l.infn.it/p3>