LIS Nepal

Lalitpur, Nepal

EFK Stack: Documentation

Documented by

Aayam Adhikari
Software Intern - 2024

1. What is the EFK stack?

EFK, which stands for Elasticsearch, Fluentd, and Kibana, is a popular open-source stack used for collecting, processing, and visualizing logs.

Elasticsearch is a distributed search and analytics engine that can store and index large amounts of data. Fluentd is a data collector that can gather data from a variety of sources and send it to Elasticsearch for indexing. Kibana is a data visualization tool that allows users to interact with data stored in Elasticsearch and create custom dashboards and visualizations. Together, these three components provide a comprehensive logging solution that can handle a large amount of data and provide real-time insights into the behavior of complex systems. Below is a brief intro to each of these:

Elasticsearch: Elasticsearch is a distributed search and analytics engine designed for storing and indexing large volumes of data. It provides fast search capabilities and is highly scalable, making it ideal for use cases such as logging, search, and analytics within the EFK stack.

Fluentd: Fluentd is a versatile data collector that gathers logs from various sources and sends them to Elasticsearch for indexing. It supports a wide range of inputs and outputs, processes logs in real-time, and can enrich logs with metadata before sending them to Elasticsearch.

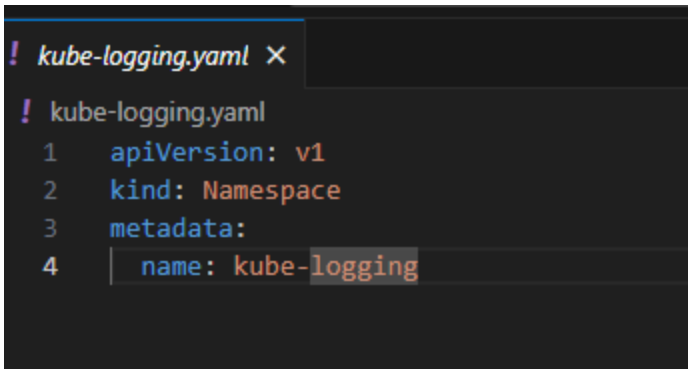
Kibana: Kibana is a powerful data visualization tool that allows users to interact with data stored in Elasticsearch. It enables users to create custom dashboards, charts, and visualizations to analyze and monitor logs in real-time. Kibana is essential for gaining insights into system behavior and monitoring application performance within the EFK stack.

2. Steps to set-up EFK stack for your logging solution

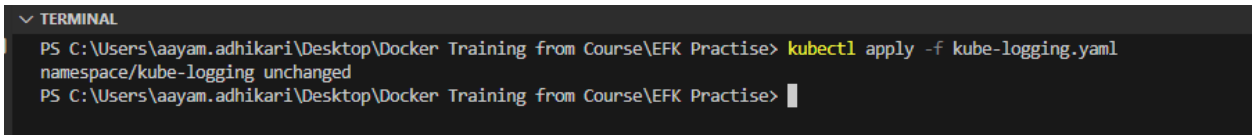
Step-1: Creating a Namespace

Before we roll out an Elasticsearch cluster, we'll first create a Namespace into which we'll install all of our logging instrumentation. Kubernetes lets you separate objects running in your cluster using a "virtual cluster" abstraction called Namespaces.

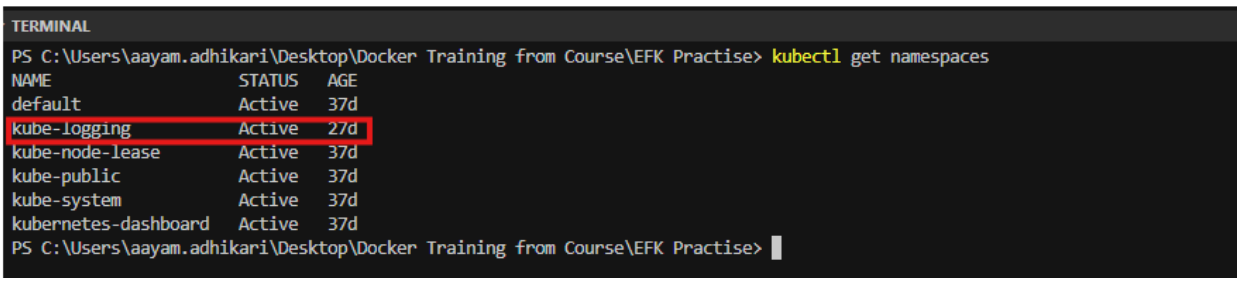
To begin, we'll create a Namespace called **kube-logging**. To do so, create a configuration file in your working directory (project folder) named as **kube-logging.yaml**, and write the following configuration code in that file:



Now, apply this configuration file using **kubectl apply -f kube-logging.yaml** in the terminal as:



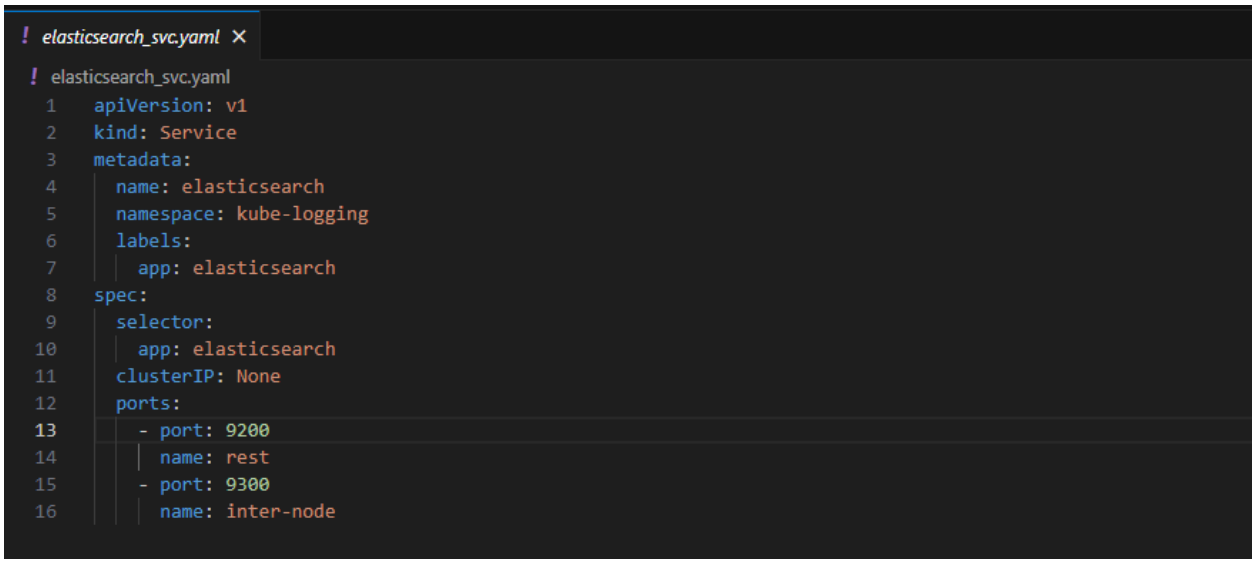
Then, confirm if the Namespace was successfully created:



Step - 2: Creating the Elasticsearch Statefulset.

To start, we'll create a headless Kubernetes service called **elasticsearch** that will define a DNS domain for the 3 Pods. A headless service does not perform load balancing or have a static IP; to learn more about headless services.

To do so, create a file named **elasticsearch_svc.yaml** and write the following kubernetes service YAML:



Create the service using **kubectl apply -f elasticsearch_svc.yaml,** and check that the service was successfully created using **kubectl get services —namespace=kube-logging** as:

```
PS C:\Users\aayam.adhikari\Desktop\Docker Training from Course\EFK Practise> kubectl get services --namespace=kube-logging
NAME            TYPE        CLUSTER-IP      EXTERNAL-IP   PORT(S)             AGE
elasticsearch   ClusterIP   None            <none>        9200/TCP,9300/TCP   27d
```

Now, create the Statefulset:

A Kubernetes StatefulSet allows you to assign a stable identity to Pods and grant them stable, persistent storage. Elasticsearch requires stable storage to persist data across Pod rescheduling and restarts.

First, open a file named **elasticsearch_stetefulset.yaml,** and write the following kubernetes yaml configuration in it:

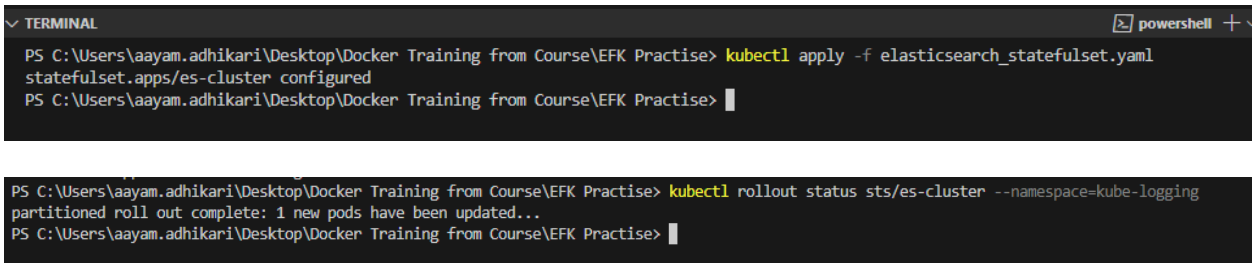
elasticsearch_statefulset.yaml

```yaml
apiVersion: apps/v1
kind: StatefulSet
metadata:
  name: es-cluster
  namespace: kube-logging
spec:
  serviceName: elasticsearch
  replicas: 1
  selector:
    matchLabels:
      app: elasticsearch
  template:
    metadata:
      labels:
        app: elasticsearch
    spec:
      containers:
      - name: elasticsearch
        image: docker.elastic.co/elasticsearch/elasticsearch:7.5.0
        resources:
            limits:
              cpu: 1000m
            requests:
              cpu: 100m
        ports:
        - containerPort: 9200
          name: rest
          protocol: TCP
        - containerPort: 9300
          name: inter-node
          protocol: TCP
```

```yaml
        volumeMounts:
        - name: data
          mountPath: /usr/share/elasticsearch/data
        env:
          - name: cluster.name
            value: k8s-logs
          - name: node.name
            valueFrom:
              fieldRef:
                fieldPath: metadata.name
          - name: discovery.seed_hosts
            value: "es-cluster-0.elasticsearch,es-cluster-1.elasticsearch,es-cluster-2.elasticsearch"
          - name: cluster.initial_master_nodes
            value: "es-cluster-0"
          - name: ES_JAVA_OPTS
            value: "-Xms512m -Xmx512m"
      initContainers:
      - name: fix-permissions
        image: busybox
        command: ["sh", "-c", "chown -R 1000:1000 /usr/share/elasticsearch/data"]
        securityContext:
          privileged: true
        volumeMounts:
        - name: data
          mountPath: /usr/share/elasticsearch/data
      - name: increase-vm-max-map
        image: busybox
```

```
      - name: increase-vm-max-map
        image: busybox
        command: ["sysctl", "-w", "vm.max_map_count=262144"]
        securityContext:
          privileged: true
      - name: increase-fd-ulimit
        image: busybox
        command: ["sh", "-c", "ulimit -n 65536"]
        securityContext:
          privileged: true
  volumeClaimTemplates:
  - metadata:
      name: data
      labels:
        app: elasticsearch
    spec:
      accessModes: [ "ReadWriteOnce" ]
      # storageClassName: ""
      resources:
        requests:
          storage: 3Gi
```

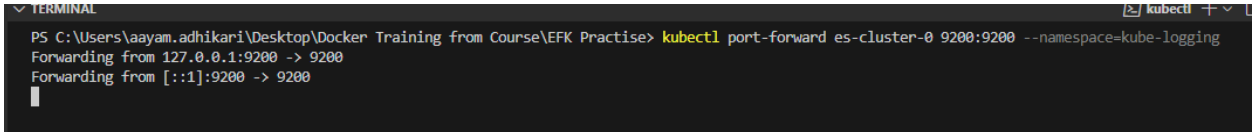
Now, deploy the stateful set using command **kubectl apply -f elasticsearch_statefulset.yaml** and monitor the StatefulSet as it is rolled out using **kubectl rollout status** as follows:



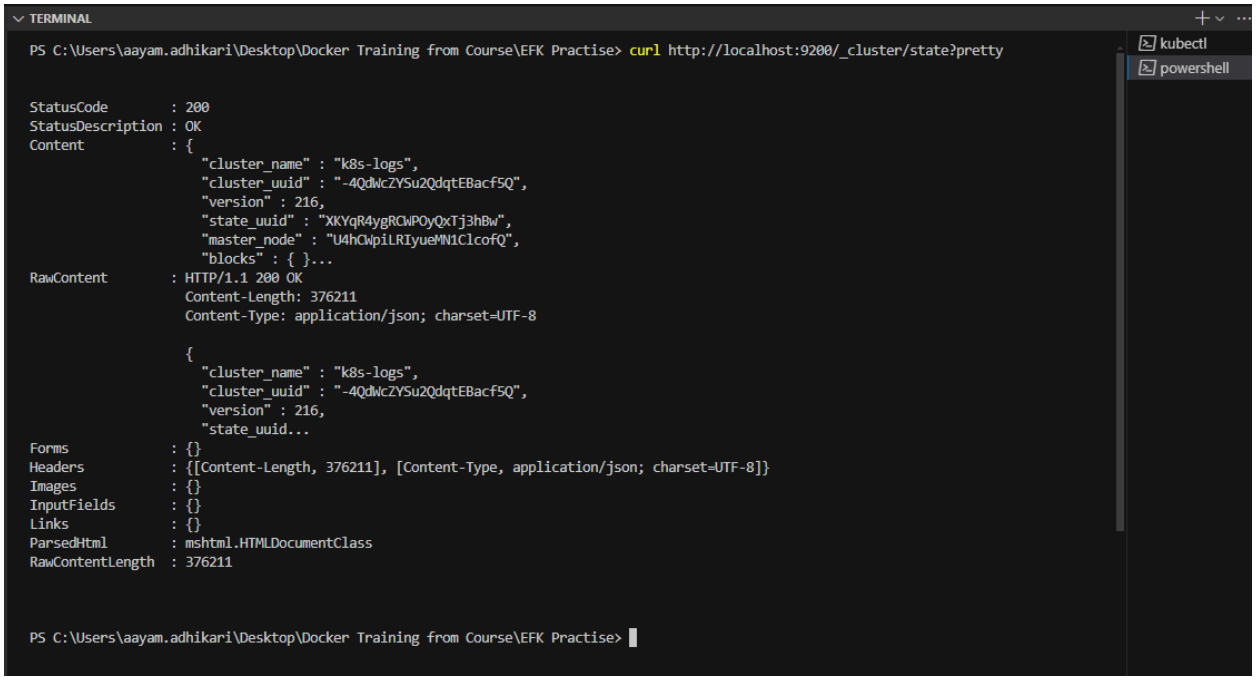
Here, I've only deployed 1 pod, due to the memory limitation and also, because the minikube cluster only includes 1 node (which acts as both the master node and worker node).

Once the Pods have been deployed, you can check that your Elasticsearch cluster is functioning correctly by performing a request against the REST API.

To do so, first forward the local port 9200 to the port 9200 on one of the Elasticsearch nodes (es-cluster-0) using kubectl port-forward:



Now, In a separate terminal window, perform a curl request against the REST API as:



Step - 3: Creating the Kibana Deployment and Service.

To launch Kibana on Kubernetes, we'll create a Service called **kibana**, and a Deployment consisting of one Pod replica. You can scale the number of replicas depending on your production needs, and optionally specify a **LoadBalancer** type for the Service to load balance requests across the Deployment pods.

This time, we'll create the Service and Deployment in the same file named **kibana.yaml** as:

kibana.yaml

```yaml
apiVersion: v1
kind: Service
metadata:
  name: kibana
  namespace: kube-logging
  labels:
    app: kibana
spec:
  ports:
  - port: 5601
  selector:
    app: kibana
---
apiVersion: apps/v1
kind: Deployment
metadata:
  name: kibana
  namespace: kube-logging
  labels:
    app: kibana
spec:
  replicas: 1
  selector:
    matchLabels:
      app: kibana
  template:
    metadata:
      labels:
        app: kibana
```

```yaml
  template:
    metadata:
      labels:
        app: kibana
    spec:
      containers:
      - name: kibana
        image: docker.elastic.co/kibana/kibana:7.2.0
        resources:
          limits:
            cpu: 1000m
          requests:
            cpu: 100m
        env:
          - name: ELASTICSEARCH_URL
            value: http://elasticsearch:9200
        ports:
        - containerPort: 5601
```

Now, roll out the Service and Deployment using **kubectl** as follows:

```
TERMINAL
PS C:\Users\aayam.adhikari\Desktop\Docker Training from Course\EFK Practise> kubectl apply -f kibana.yaml
service/kibana unchanged
deployment.apps/kibana configured
PS C:\Users\aayam.adhikari\Desktop\Docker Training from Course\EFK Practise>
```

```
PS C:\Users\aayam.adhikari\Desktop\Docker Training from Course\EFK Practise> kubectl rollout status deployment/kibana --namespace=kube-logging
deployment "kibana" successfully rolled out
PS C:\Users\aayam.adhikari\Desktop\Docker Training from Course\EFK Practise>
```

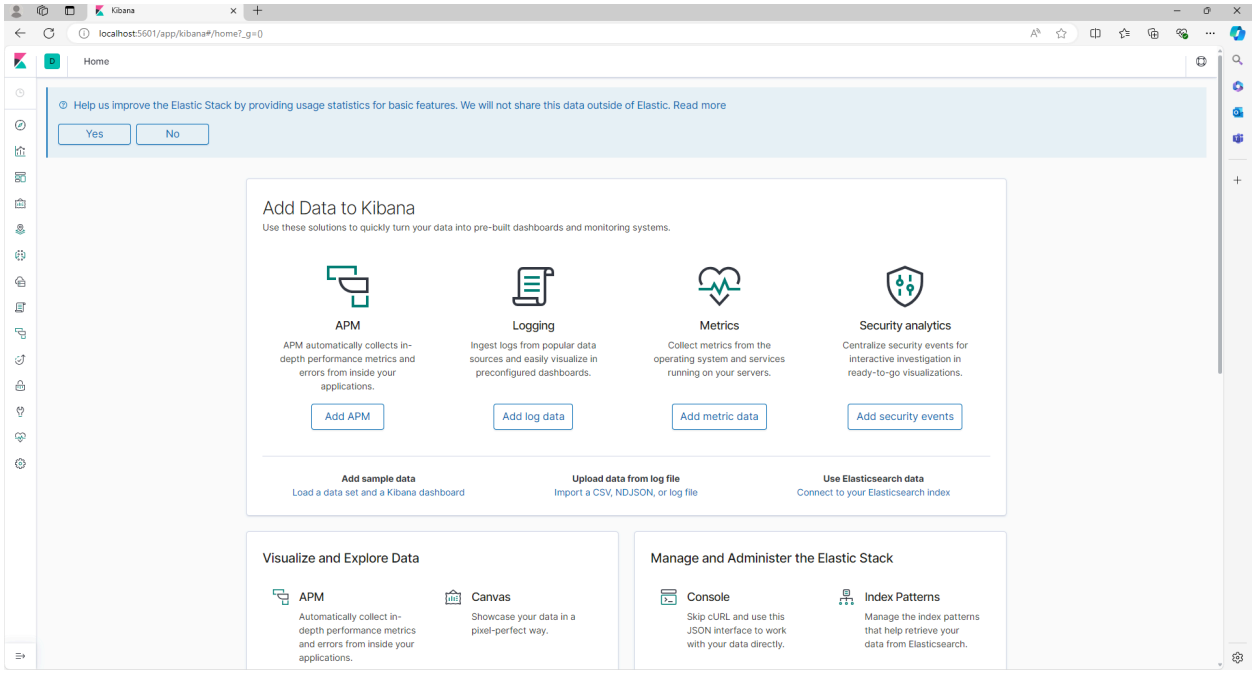
Now, To access the Kibana interface, we'll once again forward a local port to the Kubernetes node running Kibana. Grab the Kibana Pod details using **kubectl get** as and also forward the port **5601** to port **5601** on this pod:

```
TERMINAL

PS C:\Users\aayam.adhikari\Desktop\Docker Training from Course\EFK Practise> kubectl get pods --namespace=kube-logging
NAME                    READY   STATUS    RESTARTS      AGE
es-cluster-0            1/1     Running   5 (16m ago)   28d
fluentd-7rz9g           1/1     Running   5 (16m ago)   28d
kibana-bc5c4875-bgkqc   1/1     Running   5 (16m ago)   28d
PS C:\Users\aayam.adhikari\Desktop\Docker Training from Course\EFK Practise> kubectl port-forward kibana-bc5c4875-bgkqc 5601:5601 --namespace=kube-logging
Forwarding from 127.0.0.1:5601 -> 5601
Forwarding from [::1]:5601 -> 5601
```

Then, In your web browser, visit the following URL:

http://localhost:5601

If you see the following Kibana welcome page, you've successfully deployed Kibana into your Kubernetes cluster:

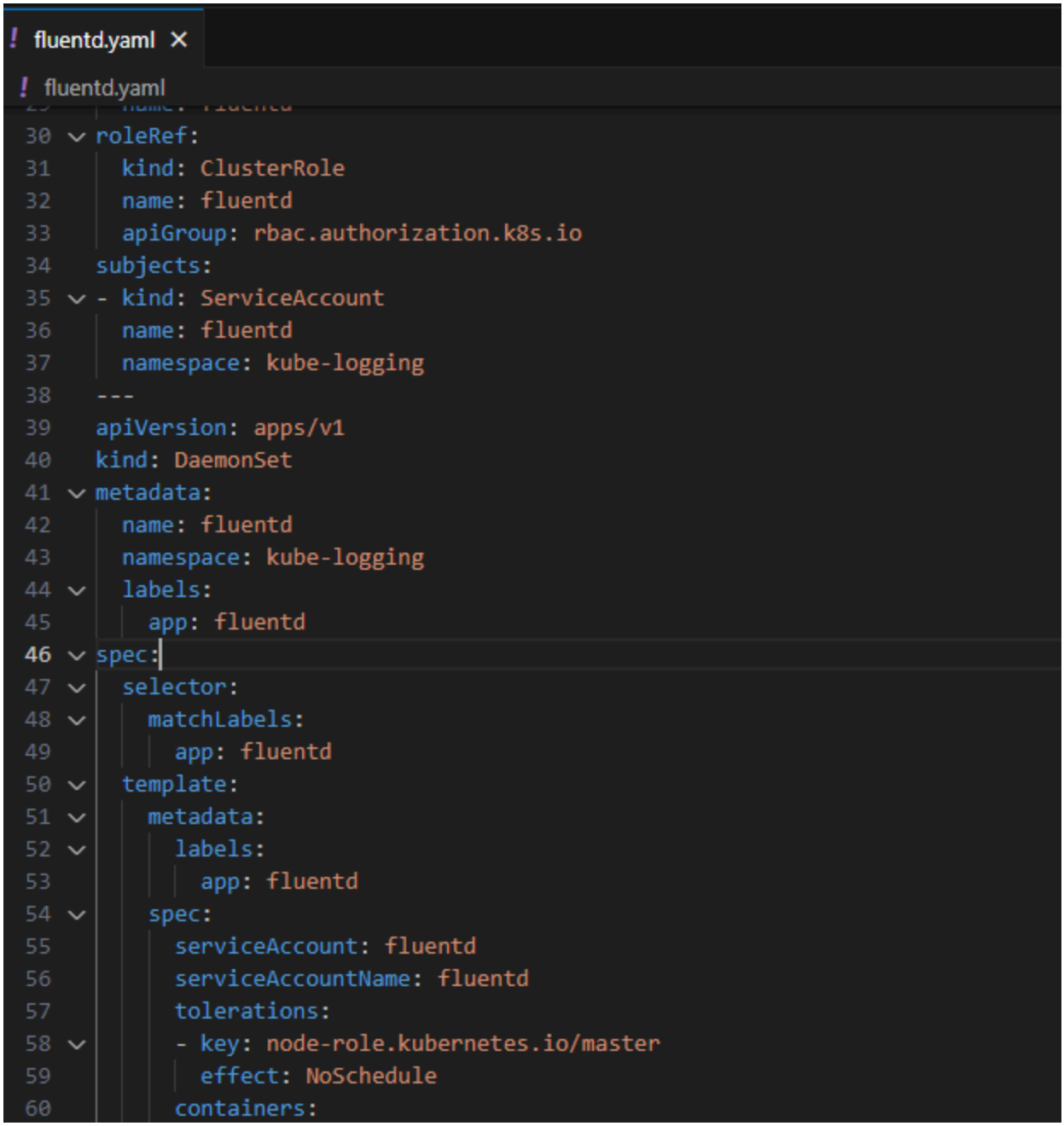


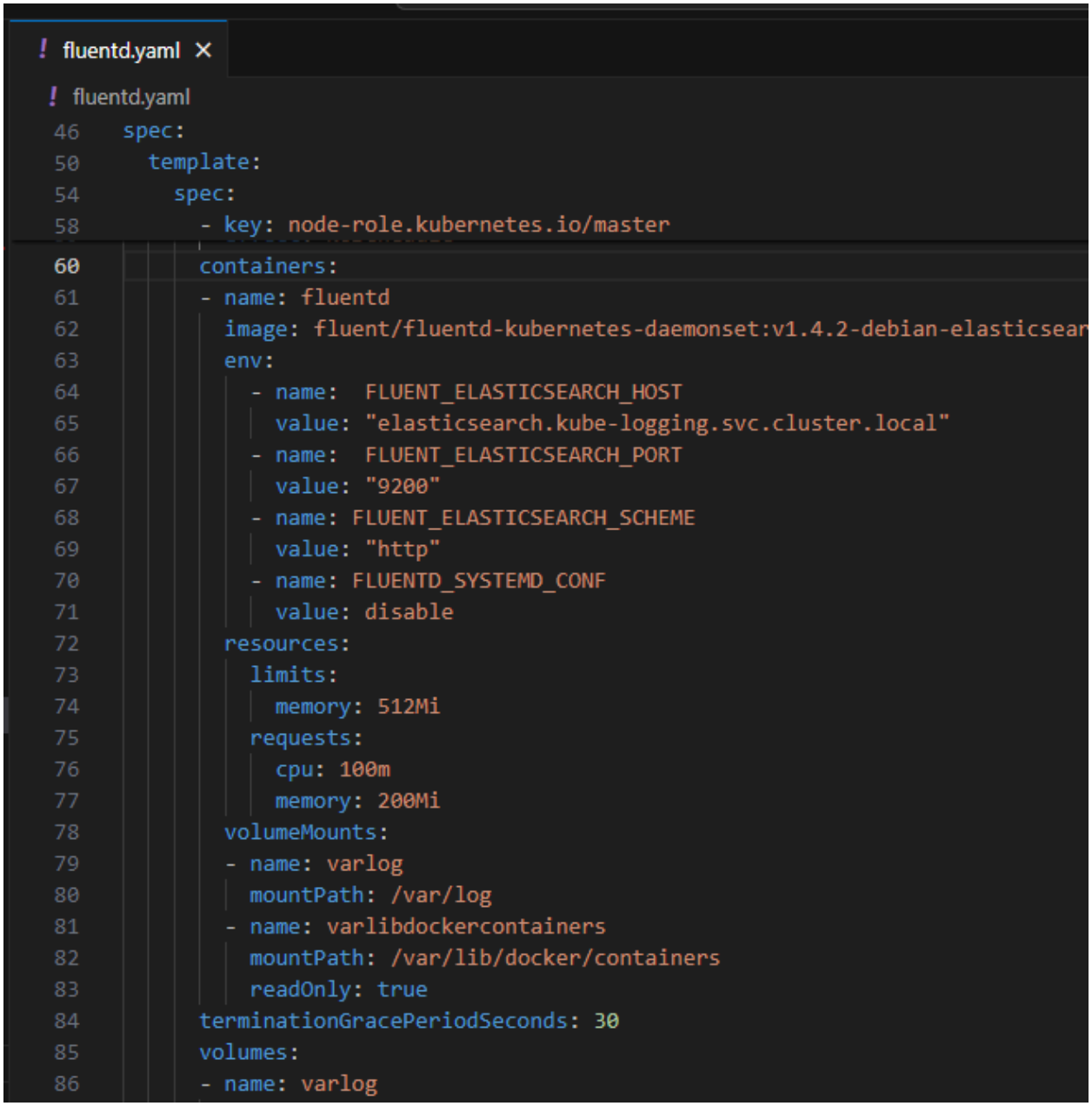
Step - 4: Creating the Fluentd DaemonSet.

Create the file named **fluentd.yaml** and write the following configuration as:

fluentd.yaml

```yaml
apiVersion: v1
kind: ServiceAccount
metadata:
  name: fluentd
  namespace: kube-logging
  labels:
    app: fluentd
---
apiVersion: rbac.authorization.k8s.io/v1
kind: ClusterRole
metadata:
  name: fluentd
  labels:
    app: fluentd
rules:
- apiGroups:
  - ""
  resources:
  - pods
  - namespaces
  verbs:
  - get
  - list
  - watch
---
kind: ClusterRoleBinding
apiVersion: rbac.authorization.k8s.io/v1
metadata:
  name: fluentd
roleRef:
  kind: ClusterRole
  name: fluentd
```


fluentd.yaml
fluentd.yaml
roleRef:
kind: ClusterRole
name: fluentd
apiGroup: rbac.authorization.k8s.io
subjects:
- kind: ServiceAccount
name: fluentd
namespace: kube-logging
---
apiVersion: apps/v1
kind: DaemonSet
metadata:
name: fluentd
namespace: kube-logging
labels:
app: fluentd
spec:
selector:
matchLabels:
app: fluentd
template:
metadata:
labels:
app: fluentd
spec:
serviceAccount: fluentd
serviceAccountName: fluentd
tolerations:
- key: node-role.kubernetes.io/master
effect: NoSchedule
containers:


fluentd.yaml
fluentd.yaml
46 spec:
50 template:
54 spec:
58 - key: node-role.kubernetes.io/master
60 containers:
61 - name: fluentd
62 image: fluent/fluentd-kubernetes-daemonset:v1.4.2-debian-elasticsear
63 env:
64 - name: FLUENT_ELASTICSEARCH_HOST
65 value: "elasticsearch.kube-logging.svc.cluster.local"
66 - name: FLUENT_ELASTICSEARCH_PORT
67 value: "9200"
68 - name: FLUENT_ELASTICSEARCH_SCHEME
69 value: "http"
70 - name: FLUENTD_SYSTEMD_CONF
71 value: disable
72 resources:
73 limits:
74 memory: 512Mi
75 requests:
76 cpu: 100m
77 memory: 200Mi
78 volumeMounts:
79 - name: varlog
80 mountPath: /var/log
81 - name: varlibdockercontainers
82 mountPath: /var/lib/docker/containers
83 readOnly: true
84 terminationGracePeriodSeconds: 30
85 volumes:
86 - name: varlog

```yaml
! fluentd.yaml
46    spec:
50      template:
54        spec:
61          - name: fluentd
72            resources:
74              memory: 512Mi
75            requests:
76              cpu: 100m
77              memory: 200Mi
78            volumeMounts:
79            - name: varlog
80              mountPath: /var/log
81            - name: varlibdockercontainers
82              mountPath: /var/lib/docker/containers
83              readOnly: true
84          terminationGracePeriodSeconds: 30
85          volumes:
86          - name: varlog
87            hostPath:
88              path: /var/log
89          - name: varlibdockercontainers
90            hostPath:
91              path: /var/lib/docker/containers
92
```

Now, roll out the DaemonSet using kubectl:

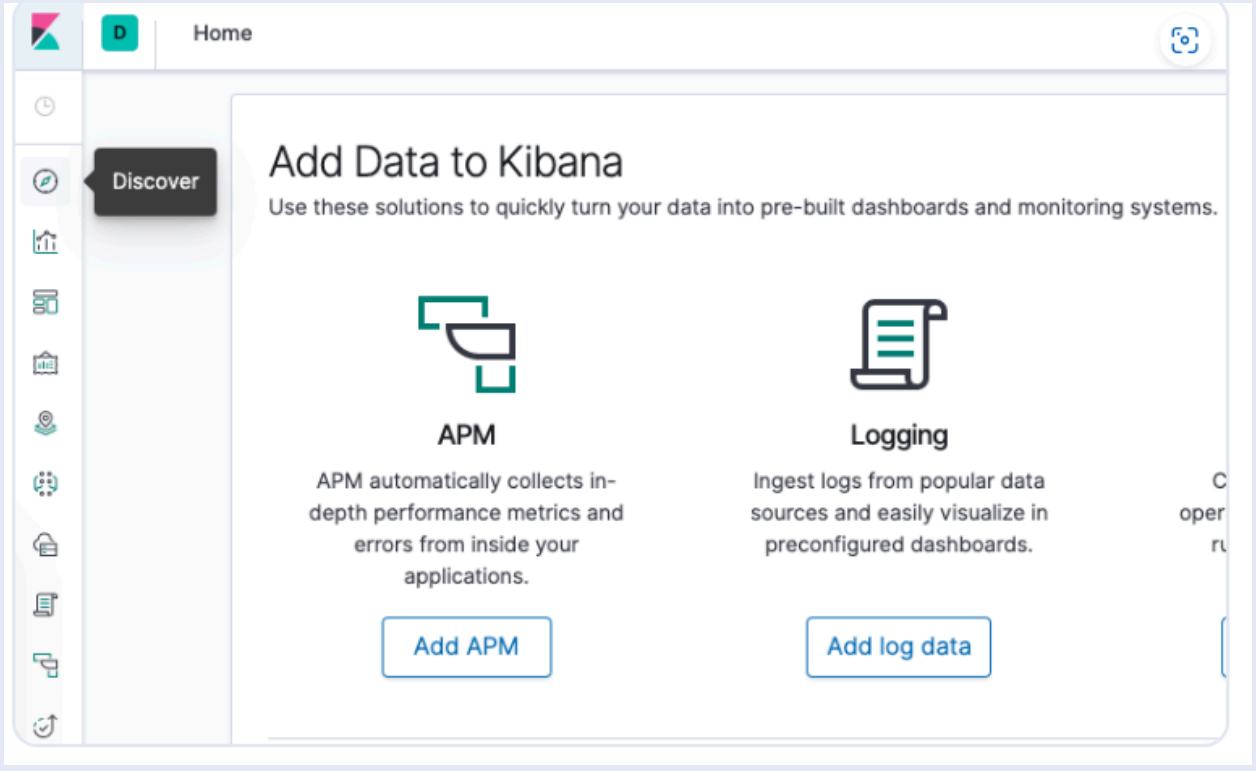
```
TERMINAL

PS C:\Users\aayam.adhikari\Desktop\Docker Training from Course\EFK Practise> kubectl apply -f fluentd.yaml
serviceaccount/fluentd unchanged
clusterrole.rbac.authorization.k8s.io/fluentd unchanged
clusterrolebinding.rbac.authorization.k8s.io/fluentd unchanged
daemonset.apps/fluentd unchanged
PS C:\Users\aayam.adhikari\Desktop\Docker Training from Course\EFK Practise>
```

Verify that your DaemonSet rolled out successfully using **kubectl** as:

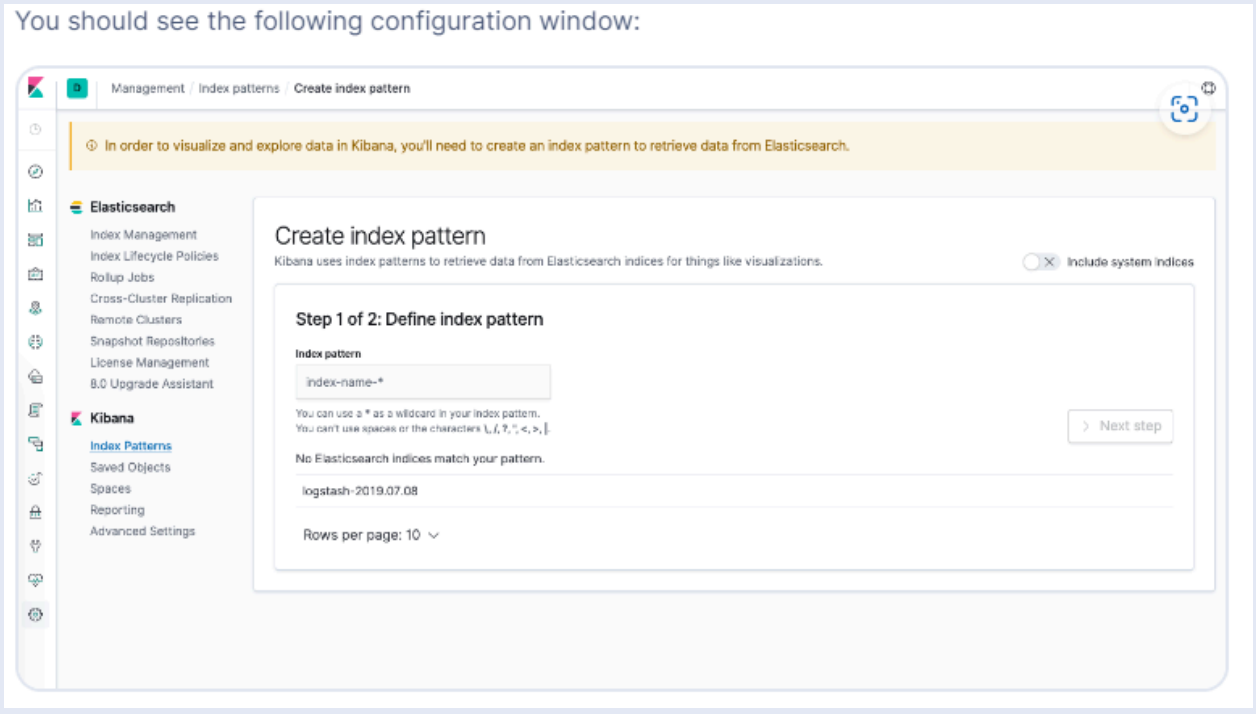
```
PS C:\Users\aayam.adhikari\Desktop\Docker Training from Course\EFK Practise> kubectl get ds --namespace=kube-logging
NAME      DESIRED   CURRENT   READY   UP-TO-DATE   AVAILABLE   NODE SELECTOR   AGE
fluentd   1         1         1       1            1           <none>          28d
PS C:\Users\aayam.adhikari\Desktop\Docker Training from Course\EFK Practise>
```

We can now check Kibana to verify that log data is being properly collected and shipped to Elasticsearch.

With the kubectl port-forward still open, navigate to http://localhost:5601 and perform the following:



You should see the following configuration window:

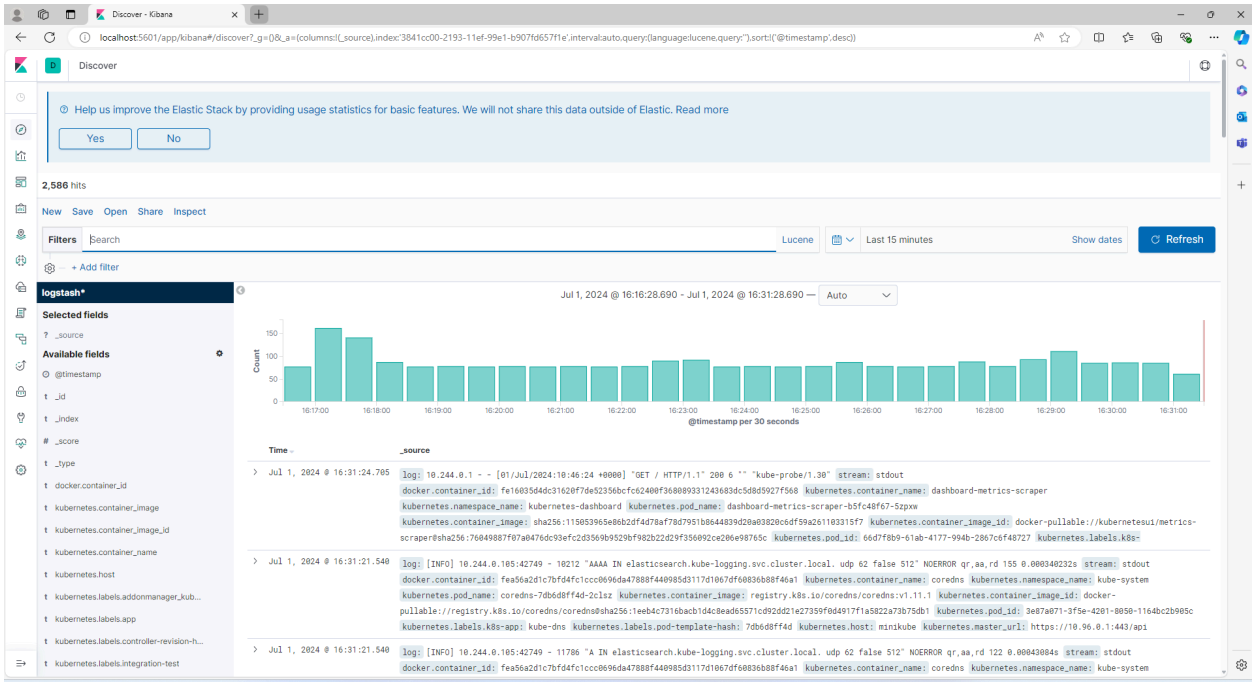


This allows you to define the Elasticsearch indices you'd like to explore in Kibana. For now, we'll just use the **logstash-*** wildcard pattern to capture all the log data in our Elasticsearch cluster. Enter **logstash-*** in the text box and click on Next step.

This allows you to configure which field Kibana will use to filter log data by time. In the dropdown, select the **@timestamp** field, and hit Create index pattern.

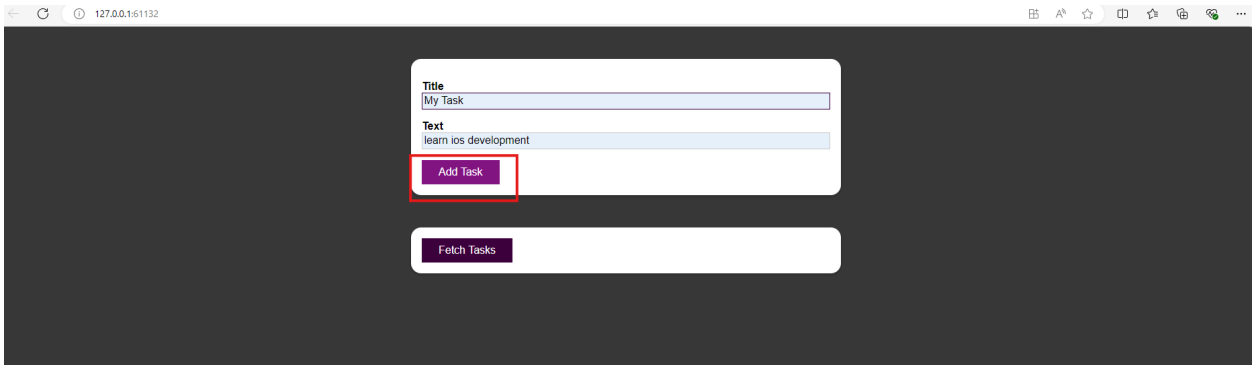
Now, hit Discover in the left hand navigation menu.

You should see a histogram graph and some recent log entries.



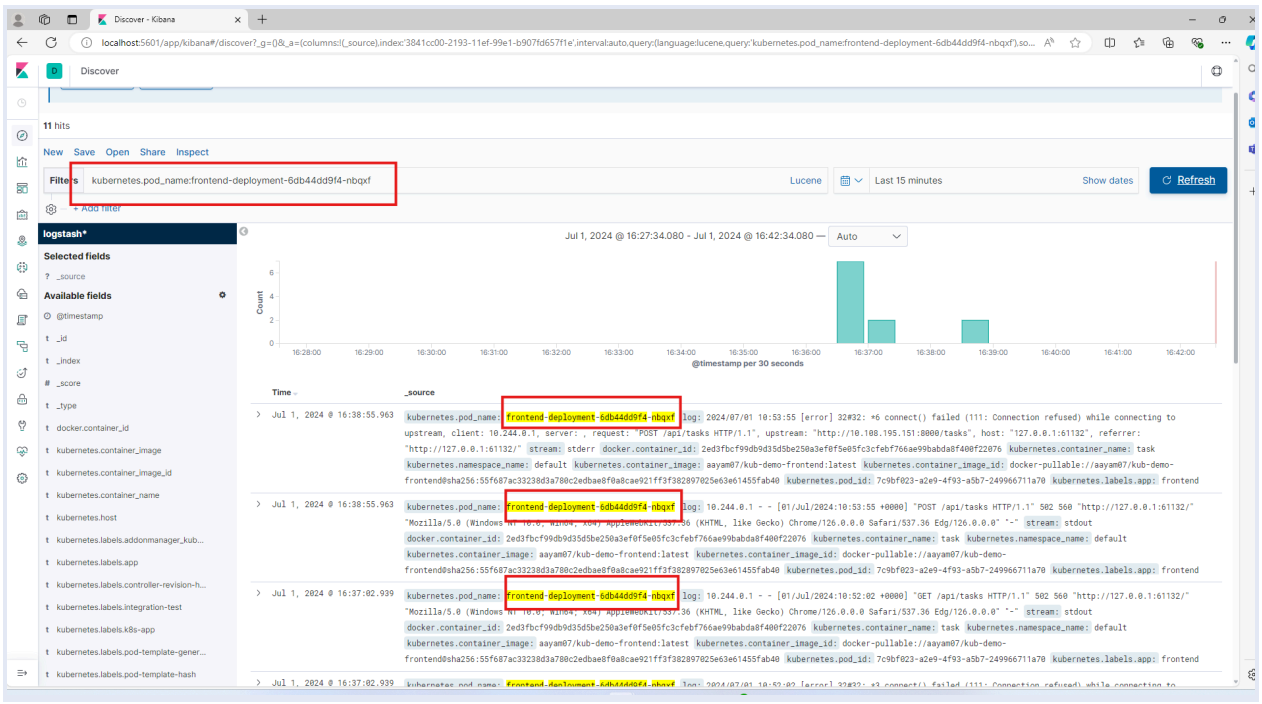
Now, for example, to see the logs of a specific **pod (or container)**, I have created a task listing project with a frontend and a backend, which is using React and Node. I have deployed this project using Kubernetes configuration files in the minikube cluster (as you can see below), and I have opened the webapp and tried to save a task as shown:

```
TERMINAL
PS C:\Users\aayam.adhikari\Desktop\Docker Training from Course\EFK Practise> kubectl get services
NAME              TYPE           CLUSTER-IP       EXTERNAL-IP   PORT(S)          AGE
auth-service      ClusterIP      10.109.119.80    <none>        3000/TCP         30d
frontend-service  LoadBalancer   10.99.125.1      <pending>     3000:32321/TCP   29d
kubernetes        ClusterIP      10.96.0.1        <none>        443/TCP          37d
task-service      LoadBalancer   10.108.195.151   <pending>     3000:31240/TCP   29d
users-service     LoadBalancer   10.106.97.9      <pending>     80:31616/TCP     31d
PS C:\Users\aayam.adhikari\Desktop\Docker Training from Course\EFK Practise> minikube service frontend-service
W0701 16:36:36.477986   11312 main.go:291] Unable to resolve the current Docker CLI context "default": context "default": context not foun
d: open C:\Users\aayam.adhikari\.docker\contexts\meta\37a8eec1ce19687d132fe29051dca629d164e2c4958ba141d5f4133a33f0688f\meta.json: The syst
em cannot find the path specified.
|-----------|-----------------|-------------|-------------------------|
| NAMESPACE |       NAME      | TARGET PORT |           URL           |
|-----------|-----------------|-------------|-------------------------|
| default   | frontend-service |        3000 | http://192.168.49.2:32321 |
|-----------|-----------------|-------------|-------------------------|
  Starting tunnel for service frontend-service.
|-----------|-----------------|-------------|-----------------------|
| NAMESPACE |       NAME      | TARGET PORT |          URL          |
|-----------|-----------------|-------------|-----------------------|
| default   | frontend-service |             | http://127.0.0.1:61132 |
|-----------|-----------------|-------------|-----------------------|
  Opening service default/frontend-service in default browser...
  ! Because you are using a Docker driver on windows, the terminal needs to be open to run it.
```



Now, navigate back to your Kibana dashboard. From the Discover page, in the search bar enter `kubernetes.pod_name:name-of-your-pod`. This filters the log data for Pods named **name-of-your-pod**. This is shown below:

```
TERMINAL
PS C:\Users\aayam.adhikari\Desktop\Docker Training from Course\EFK Practise> kubectl get pods
NAME                                  READY   STATUS    RESTARTS      AGE
auth-deployment-5d496ffd7d-2rff5      1/1     Running   3 (44m ago)   26d
frontend-deployment-6db44dd9f4-nbqxf  1/1     Running   1             171m
users-deployment-5c96898806-tn517     1/1     Running   4 (44m ago)   26d
PS C:\Users\aayam.adhikari\Desktop\Docker Training from Course\EFK Practise>
```



As you can see above, we're successfully getting logs from the pod that contains the frontend React application. Similar process can be applied for any project application.