

Fundamentals of SRE Post-mortem Template

*Adapted from the SRE Book* [*Appendix*](https://sre.google/sre-book/example-postmortem/)

# Incident Title

|  |
| --- |
| ***Metadata***  **Date:** 2024-04-13 *Do you know about* [*ISO 8601*](https://www.iso.org/iso-8601-date-and-time-format.html)*?*  **Author(s): Tanushree Aggarwal**  **Status:** *Draft |* ***Ready for Review*** *| Complete* |

## Summary

The Upcommerce deployment failed and resulted in the Upcommerce service being unavailable in the production environment due to CPU misconfiguration in the deployment file. The issue was identified and fixed, restoring service availability.

## Impact

The UpCommerce service (upcommerce-app-two) was down, customers were unable to use the service.

## Root Causes

Incorrect CPU limits configured in the deployment specification.

## Trigger

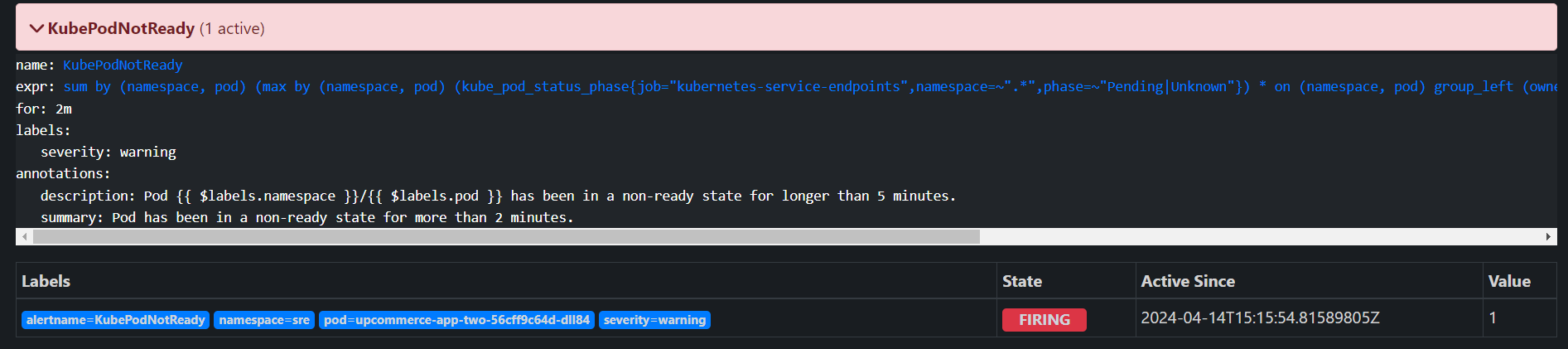
Misconfiguration in Kubernetes deployment.yml file

## Resolution

The CPU allocation for the affected pods was adjusted from 10 CPUs to the standard 1 CPU.

## Detection

* Outage was detected by Upcommerce monitoring system - Prometheus Alert Manager



* Troubleshooting via describing the deployment and the pod specs

@aggarwal-tanushree ➜ /workspaces/sre-week-two (1-week-2-project) $ **kubectl get pods -n sre**

NAME READY STATUS RESTARTS AGE

grafana-557d966c8c-t5q72 1/1 Running 0 10m

prometheus-alertmanager-0 1/1 Running 0 10m

prometheus-kube-state-metrics-65468947fb-8crhp 1/1 Running 0 10m

prometheus-prometheus-node-exporter-p7m56 1/1 Running 0 10m

prometheus-prometheus-pushgateway-76976dc66-htqw5 1/1 Running 0 10m

prometheus-server-8444b5b7f7-bb67t 2/2 Running 0 10m

**upcommerce-app-two-56cff9c64d-dll84 0/1 Pending 0 6m57s**

@aggarwal-tanushree ➜ /workspaces/sre-week-two (1-week-2-project) $ kubectl describe deployment upcommerce-app-two -n sre

Name: upcommerce-app-two

Namespace: sre

CreationTimestamp: Sun, 14 Apr 2024 15:14:33 +0000

Labels: <none>

Annotations: deployment.kubernetes.io/revision: 1

Selector: app=upcommerce-app-two

**Replicas: 1 desired | 1 updated | 1 total | 0 available | 1 unavailable**

StrategyType: RollingUpdate

MinReadySeconds: 0

RollingUpdateStrategy: 25% max unavailable, 25% max surge

Pod Template:

Labels: app=upcommerce-app-two

Containers:

upcommerce:

Image: uonyeka/upcommerce:v3

Port: 5000/TCP

Host Port: 0/TCP

Limits:

**cpu: 10**

memory: 4Gi

Environment: <none>

Mounts: <none>

Volumes: <none>

Conditions:

Type Status Reason

---- ------ ------

**Available False MinimumReplicasUnavailable**

Progressing True ReplicaSetUpdated

OldReplicaSets: <none>

NewReplicaSet: upcommerce-app-two-56cff9c64d (1/1 replicas created)

Events:

Type Reason Age From Message

---- ------ ---- ---- -------

Normal ScalingReplicaSet 2m4s deployment-controller Scaled up replica set upcommerce-app-two-56cff9c64d to 1

@aggarwal-tanushree ➜ /workspaces/sre-week-two (1-week-2-project) $

@aggarwal-tanushree ➜ /workspaces/sre-week-two (1-week-2-project) $ **kubectl describe pod upcommerce-app-two -n sre**

Name: **upcommerce-app-two-56cff9c64d-dll84**

Namespace: sre

Priority: 0

Service Account: default

Node: <none>

Labels: app=upcommerce-app-two

pod-template-hash=56cff9c64d

Annotations: <none>

**Status: Pending**

IP:

IPs: <none>

Controlled By: ReplicaSet/upcommerce-app-two-56cff9c64d

Containers:

upcommerce:

Image: uonyeka/upcommerce:v3

Port: 5000/TCP

Host Port: 0/TCP

Limits:

cpu: 10

memory: 4Gi

Requests:

cpu: 10

memory: 4Gi

Environment: <none>

Mounts:

/var/run/secrets/kubernetes.io/serviceaccount from kube-api-access-jh6gh (ro)

Conditions:

Type Status

PodScheduled False

Volumes:

kube-api-access-jh6gh:

Type: Projected (a volume that contains injected data from multiple sources)

TokenExpirationSeconds: 3607

ConfigMapName: kube-root-ca.crt

ConfigMapOptional: <nil>

DownwardAPI: true

QoS Class: Guaranteed

Node-Selectors: <none>

Tolerations: node.kubernetes.io/not-ready:NoExecute op=Exists for 300s

node.kubernetes.io/unreachable:NoExecute op=Exists for 300s

Events:

Type Reason Age From Message

---- ------ ---- ---- -------

**Warning FailedScheduling 22s (x2 over 5m46s) default-scheduler 0/1 nodes are available: 1 Insufficient cpu. preemption: 0/1 nodes are available: 1 No preemption victims found for incoming pod..**

## Action Items

1. Implement a more rigorous configuration review process for production deployments.
2. Ensure UAT testing and QA sign-off before production deployment