**DEPLOYING SPARKS ON KUBERNETES**

What is Hadoop?

Ans: **Hadoop** is an open-source software framework for storing data and running applications on clusters of commodity hardware. It provides massive storage for any kind of data, enormous processing power and the ability to handle virtually limitless concurrent tasks or jobs.

What is Spark?

Apache Spark is an open-source, distributed processing system used for big data workloads. It utilizes in-memory caching and optimized query execution for fast queries against data of any size. Simply put, Spark is a **fast and general engine for large-scale data processing**.

**PREREQUISITES:**

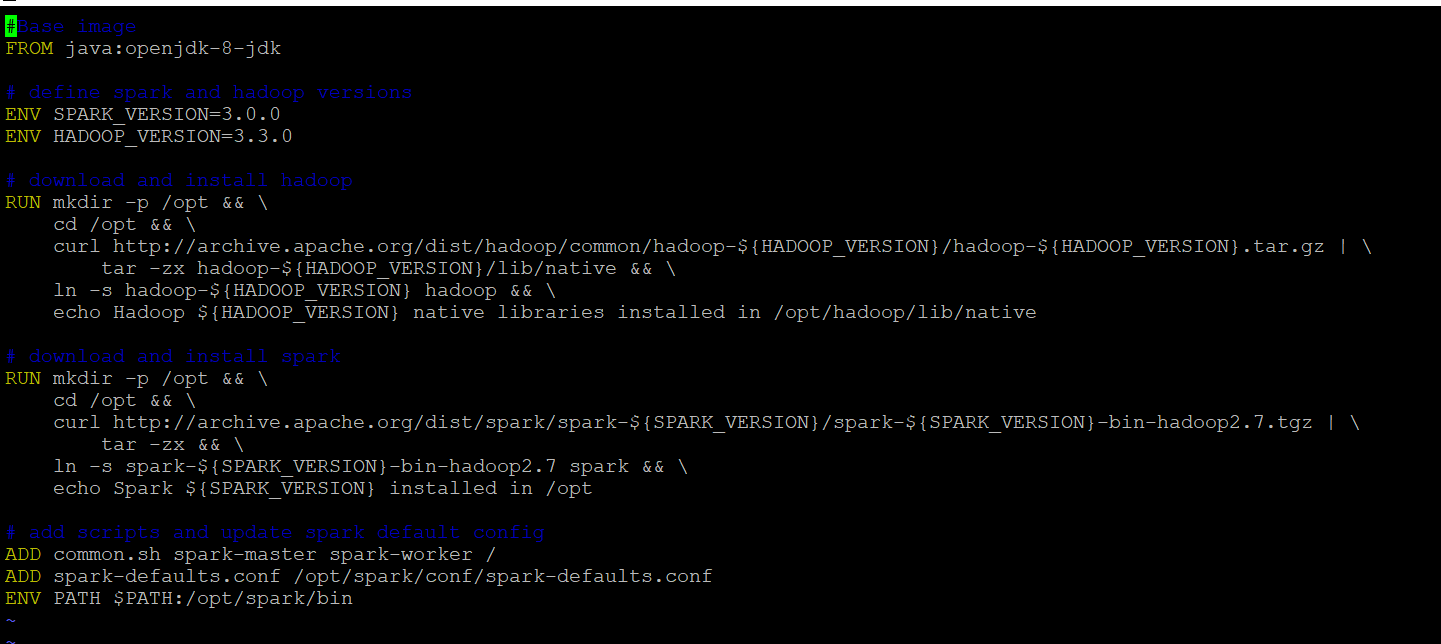
* Docker
* Kubernetes Cluster with all permissions
* Spark
* Hadoop

**PROCEDURE:**

Step 1: In our environment we have already using Docker and Kubernetes so no

Need to set up again.

Step 2: Build Docker image for Spark by writing Dockerfile.



OR

You can pull image from Docker Hub also.

Step 3: Build the Docker image by using the below command.

$ cd spark

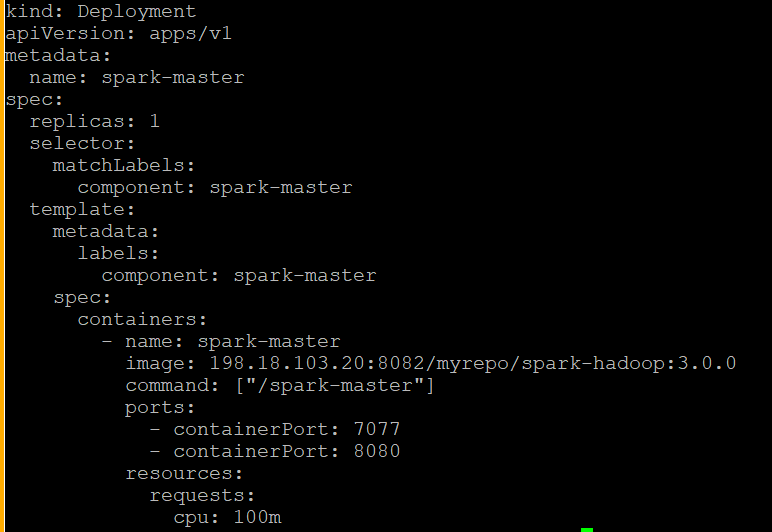
$ docker image build -t spark-hadoop:3.0.0

Step 4: Create namespace if you want, here spark-hadoop is namespace used

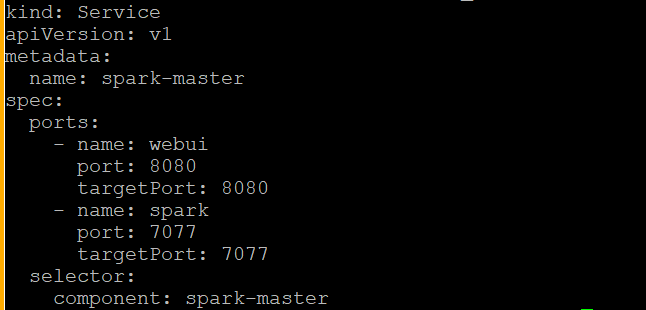
$ kubectl create namespace spark-hadoop

Step 5: Write yaml scripts for master node and Service.

spark-master-deployment.yaml:



spark-master-service.yaml:



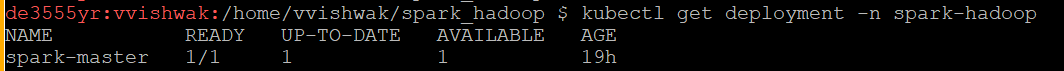
Step 6: Create the spark master deployment and start the service by running below commands.

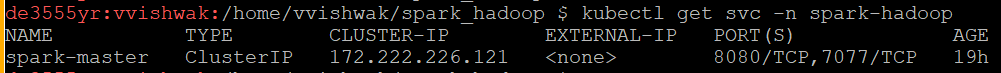
$ cd spark\_hadoop

$ kubectl create -f spark-master-deployment.yaml -n spark-hadoop

$ kubectl create -f spark-master-service.yaml –n spark-hadoop

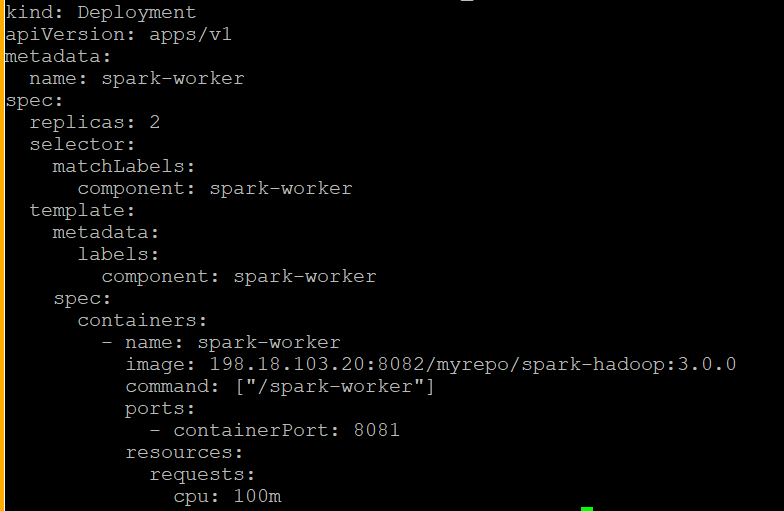
Verify:





Step 7: Write Yaml script for Worker nodes.

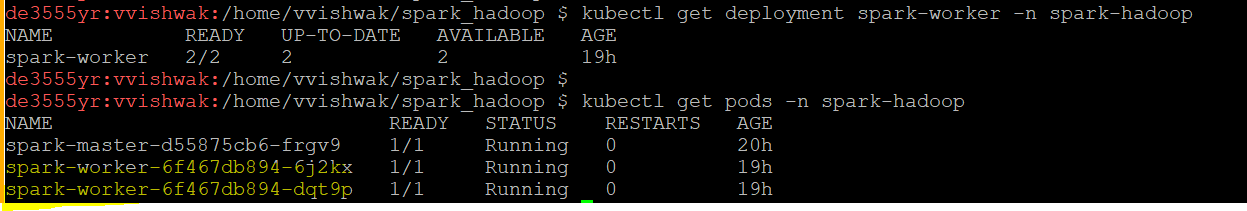
spark-worker-deployment.yaml:



Create the Spark worker Deployment:

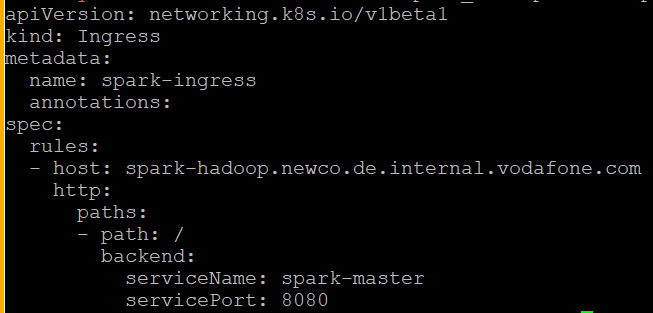
$ kubectl create -f spark-worker-deployment.yaml

Verify:



Step 8: We need to get traffic from outside will setup ingress controller.

spark-ingress.yaml



Create the Ingress object:

$ kubectl create -f spark-ingress.yaml

Test it out in the browser at <http://spark-hadoop.newco.de.internal.vodafone.com/>:

