

ASSIGNMENT 03

Sabaragamuwa University of Sri Lanka

Faculty of Computing

Department of Software Engineering

SE6103 – Parallel and Distributed Systems

Name	: K.O.K.S.Dayarathna
Reg. No	: 19APSE4299
Degree Program	: Software Engineering
Academic Period	: 3 RD Year 2 ND Semester
Due Date	: 06/01/2025

Question 1

Date:

Assignment 03

19APSE4299
K.O.K.S. Dayarathna

Question 01

- 1- When talk about architecture, Docker container share the host OS kernel, while virtual machines (VMs) include a full OS. Containers are lightweight, using fewer resources since they do not require a full OS. VMs are heavier as they include a full OS and require a hypervisor. VMs provide stronger isolation as each VM operates as a separate system. Containers share the host kernel, leading to slightly less isolation.
2. * The `-d` flag runs the container in detached mode, meaning it runs in the background.
* Why important?
Running services like Nginx in detached mode allows you to keep them running while using the terminal for each tasks.
3. * Difference between `docker run -d nginx:latest` and `docker run nginx:latest`
* `docker run -d nginx:latest` runs the Nginx container in detached mode.
* `docker run nginx:latest` runs the container in the foreground, blocking the terminal until stopped.
* For long-running services like Nginx, you would use `docker run -d nginx:latest` to keep the service running in the background.

4. Purpose → Maps port 8080 on the host to port 80 in the container.

Why necessary → It allows the host system to access services running inside the container through the specified port

5. Hadoop is an open-source framework for storing and processing large datasets across a distributed network of computers.

It is used for big data analytics, handling tasks ~~that~~ like data storage, processing and analysis.

6. When talk the speed, Spark processes data in-memory making it much faster than Hadoop's disk-based MapReduce. As well as ~~Hadoop~~ Hadoop provides APIs in multiple languages (Python, Scala, Java, R). Spark can run on Hadoop, standalone clusters, or in the cloud.

Question 2

```
PS D:\Academic\Semester 6\Docker\Assignment 3> docker-compose up -d
time="2025-01-06T14:12:25+05:30" level=warning msg="D:\Academic\Semester 6\Docker\Assignment 3\docker-compose.yml: the attribute `version` is obsolete, it will be ignored, please remove it to avoid potential confusion"
[+] Running 3/3
 ✓ Container historyserver Started 6.1s
 ✓ Container namenode Started 4.4s
 ✓ Container datanode Started 0.6s
```

```
PS D:\Academic\Semester 6\Docker\Assignment 3> docker exec -it namenode hdfs dfs -put /sample.txt /input
2025-01-06 08:58:20,929 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false

What's next:
Try Docker Debug for seamless, persistent debugging tools in any container or image → docker debug namenode
Learn more at https://docs.docker.com/go/debug-cli/
```

```
PS D:\Academic\Semester 6\Docker\Assignment 3> docker exec -it namenode hdfs dfs -ls /input
-rw-r--r-- 3 root supergroup 181 2025-01-06 08:58 /input
```

```
What's next:
Try Docker Debug for seamless, persistent debugging tools in any container or image → docker debug namenode
Learn more at https://docs.docker.com/go/debug-cli/
```

```
PS D:\Academic\Semester 6\Docker\Assignment 3> docker cp /path/to/wordcount.jar namenode:/opt/hadoop-3.2.1/wordcount.jar
CreateFile D:\path: The system cannot find the file specified.
```

```
2025-01-06 09:09:43,141 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
Hadoop 2
a 2
about 1
across 1
amounts 1
an 1
and 1
computers. 1
data 1
framework 2
interesting 1
is 2
large 1
learn 1
network 1
networking. 1
of 2
open-source 1
processes 1
stores 1
that 1
to 1
very 1
```