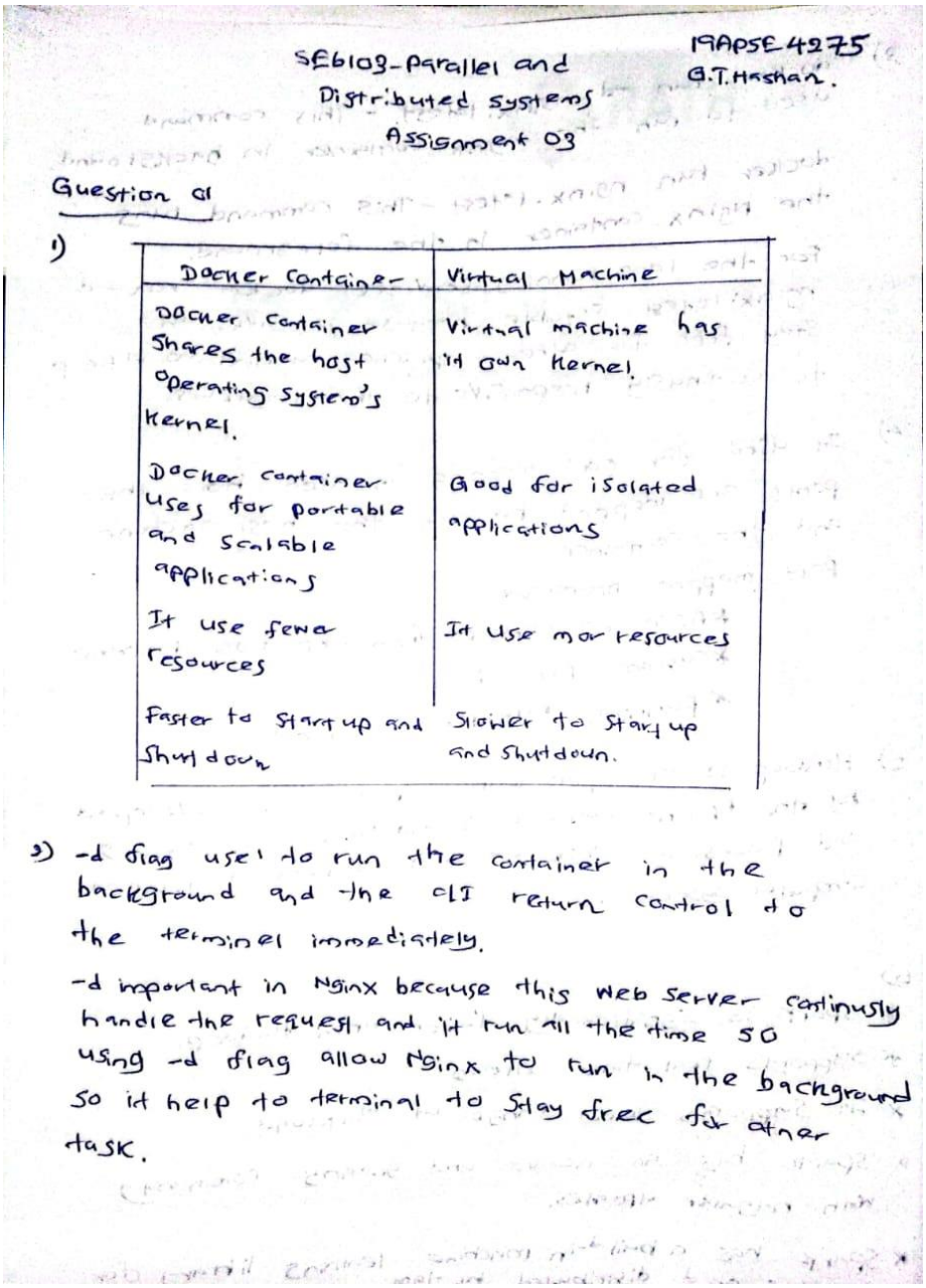
Faculty of Computing

Sabaragamuwa University of Sri Lanka

SE6103- Parallel and Distributed Systems Assignment 03

19APSE4275

Question 1



SE6103-Parallel and Distributed systems
Assignment 03

19APSE4275
G.T.Hashan.

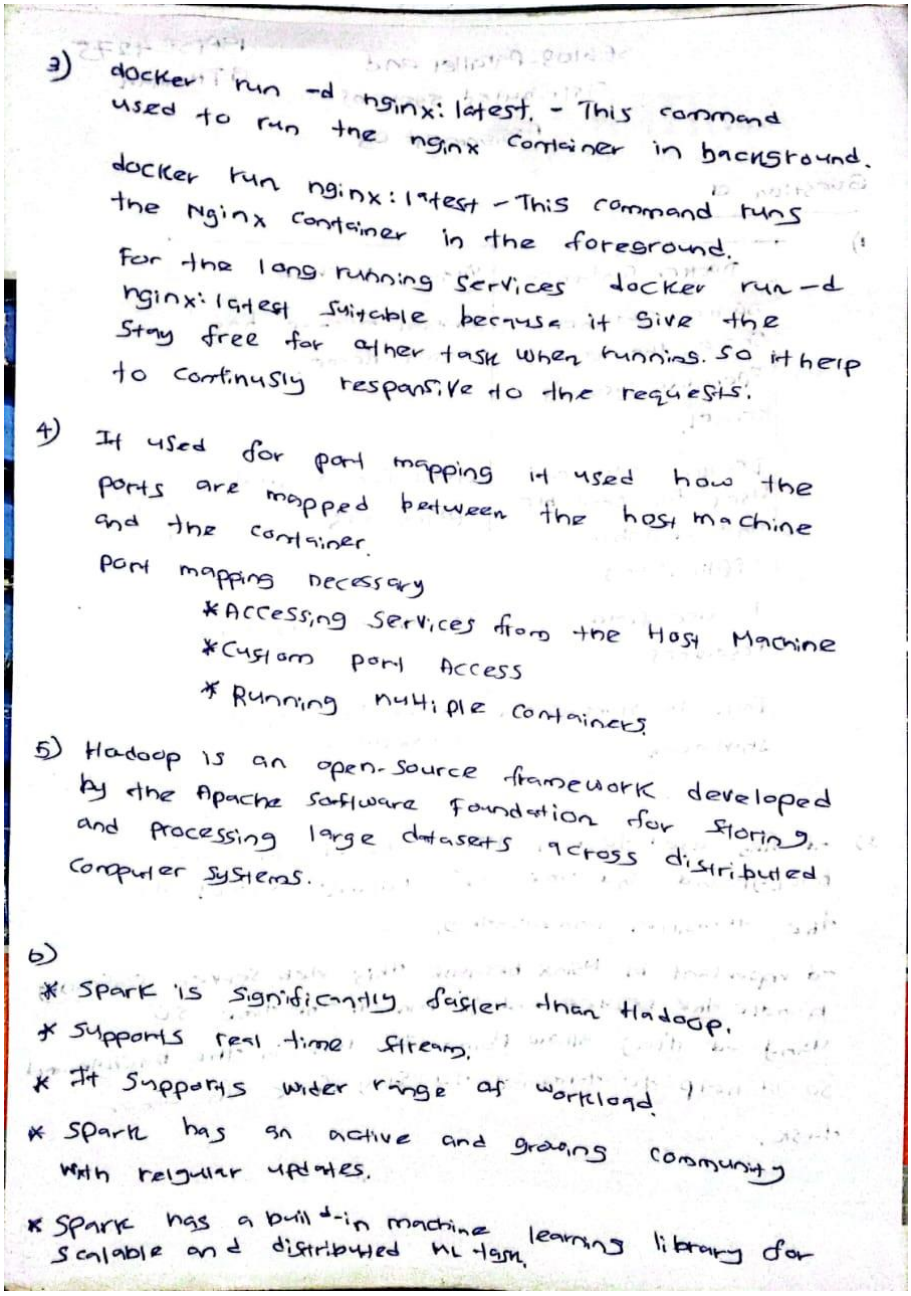
Question 01

1)

Docker Container	Virtual Machine
Docker container shares the host operating system's kernel.	Virtual machine has it own kernel.
Docker container uses for portable and scalable applications	Good for isolated applications
It use fewa resources	It use mor resources
Faster to start up and shut down	Slower to start up and shutdown.

2) -d diag use to run the container in the background and the CLI return control to the terminal immediately.

-d important in Nginx because this web server continusly handle the request, and it run all the time so using -d diag allow Nginx to run in the background so it help to terminal to stay free for other task.



3) docker run -d nginx:latest. - This command used to run the nginx container in background.

docker run nginx:latest - This command runs the Nginx container in the foreground.

For the long running services docker run -d nginx:latest suitable because it give the stay free for other task when running. So it help to continusly responsive to the requests.

4) It used for port mapping it used how the ports are mapped between the host machine and the container.

Port mapping necessary

* Accessing services from the Host Machine
* Custom port Access
* Running multiple containers

5) Hadoop is an open-source framework developed by the Apache Software Foundation for storing and processing large datasets across distributed computer systems.

6)

* Spark is significantly faster than Hadoop.
* Supports real time stream.
* It supports wider range of workload.
* Spark has an active and growing community with regular updates.
* Spark has a built-in machine learning library for scalable and distributed ML tasks.

Question 2

docker-compose up –d

```
PS C:\sabaragamuwa 6th\ParallelANDDistributedSystemAssigment3> docker-compose up -d
time="2025-01-06T14:22:44+05:30" level=warning msg="C:\\sabaragamuwa 6th\\ParallelANDDistributedSystemAssigment3\\docker-compose.yml: the attribute `version` is obsolete, it will
 be ignored, please remove it to avoid potential confusion"
[+] Running 22/3
 ✓ namenode Pulled                                                                                                                                                    209.8s
 ✓ datanode Pulled                                                                                                                                                    209.8s
 ✓ historyserver Pulled                                                                                                                                               209.8s
[+] Running 6/6
 ✓ Network parallelanddistributedsystemassigment3_default                 Created                                                                                      0.2s
 ✓ Volume "parallelanddistributedsystemassigment3_datanode-data"          Created                                                                                      0.0s
 ✓ Volume "parallelanddistributedsystemassigment3_namenode-data"          Created                                                                                      0.0s
 ✓ Container namenode                                                     Started                                                                                      4.3s
 ✓ Container datanode                                                     Started                                                                                      4.1s
 ✓ Container historyserver                                                Started                                                                                      4.6s
```

docker exec -it namenode hdfs dfs -mkdir -p /input

```
PS C:\sabaragamuwa 6th\ParallelANDDistributedSystemAssigment3> docker exec -it namenode hdfs dfs -mkdir -p /input

What's next:
    Try Docker Debug for seamless, persistent debugging tools in any container or image → docker debug namenode
```

docker cp ./sample.txt namenode:/tmp/sample.txt

```
PS C:\sabaragamuwa 6th\ParallelANDDistributedSystemAssigment3> docker cp ./sample.txt namenode:/tmp/sample.txt
Successfully copied 2.05kB to namenode:/tmp/sample.txt
```

docker exec -it namenode hdfs dfs -put /tmp/sample.txt /input

```
PS C:\sabaragamuwa 6th\ParallelANDDistributedSystemAssigment3> docker exec -it namenode hdfs dfs -put /tmp/sample.txt /input
2025-01-06 09:14:49,420 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 C:\sabaragamuwa 6th\ParallelANDDistributedSystemAssigment3> docker exec -it namenode hdfs dfs -ls /input
Found 1 items
-rw-r--r--   3 root supergroup        181 2025-01-06 09:14 /input/sample.txt
```

docker exec -it namenode hadoop jar /opt/hadoop-3.2.1/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.2.1.jar wordcount /input /output

```
PS C:\sabaragamuwa 6th\ParallelANDDistributedSystemAssigment3> docker exec -it namenode hadoop jar /opt/hadoop-3.2.1/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.2.1.jar wo
rdcount /input /output
2025-01-06 09:25:05,843 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-01-06 09:25:05,977 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-01-06 09:25:05,977 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2025-01-06 09:25:06,446 INFO input.FileInputFormat: Total input files to process : 1
2025-01-06 09:25:06,478 INFO mapreduce.JobSubmitter: number of splits:1
2025-01-06 09:25:06,703 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local2114181928_0001
2025-01-06 09:25:06,703 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-01-06 09:25:06,817 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2025-01-06 09:25:06,818 INFO mapreduce.Job: Running job: job_local2114181928_0001
2025-01-06 09:25:06,821 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2025-01-06 09:25:06,828 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-01-06 09:25:06,828 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-01-06 09:25:06,829 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
2025-01-06 09:25:06,922 INFO mapred.LocalJobRunner: Waiting for map tasks
2025-01-06 09:25:06,924 INFO mapred.LocalJobRunner: Starting task: attempt_local2114181928_0001_m_000000_0
2025-01-06 09:25:06,956 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2025-01-06 09:25:06,956 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2025-01-06 09:25:07,011 INFO mapred.Task:  Using ResourceCalculatorProcessTree : [ ]
2025-01-06 09:25:07,015 INFO mapred.MapTask: Processing split: hdfs://namenode:8020/input/sample.txt:0+181
2025-01-06 09:25:07,089 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
```

docker exec -it namenode hdfs dfs -cat /output/part-r-00000

```
PS C:\sabaragamuwa 6th\ParallelANDDistributedSystemAssigment3> docker exec -it namenode hdfs dfs -cat /output/part-r-00000
2025-01-06 09:25:53,011 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
```

```
processes       1
stores  1
that    1
to      1
very    1

What's next:
    Try Docker Debug for seamless, persistent debugging tools in any container or image → docker debug namenode
```