



ESCUELA POLITÉCNICA NACIONAL
ESCUELA DE FORMACIÓN DE TECNÓLOGOS

APLICACIONES DISTRIBUIDAS (TDSD523)



ASIGNATURA:

Aplicaciones Distribuidas

PROFESOR:

Ing. Sergio Granizo MSc.

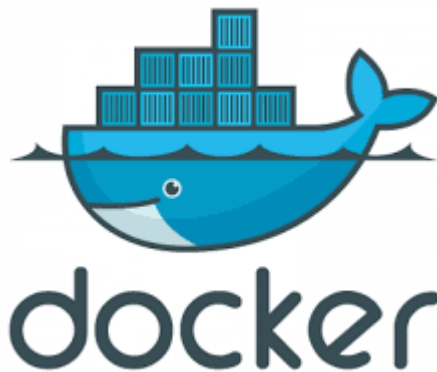
PERÍODO ACADÉMICO:

2025-A

TALLER 4 – 2B

TÍTULO:

MAP-REDUCE EN DOCKER



Nombre del Estudiante:

Guerra Lovato Josué Eduard

PROPÓSITO DE LA PRÁCTICA

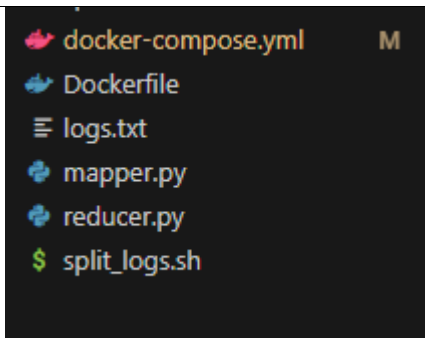
- Comprender los conceptos fundamentales de map reduce.
- Realizar un ejercicio evidenciando su funcionamiento

ANÁLISIS DE ACCESOS FUERA DEL HORARIO LABORAL CON MAPREDUCE

- Aplicar el modelo MapReduce para detectar y contar accesos al sistema que ocurren fuera del horario laboral, definidos entre las 08:00:00 y las 18:00:00.
- El alumno debe identificar qué usuarios accedieron fuera del horario establecido, y cuántas veces lo hicieron, utilizando un enfoque distribuido con 4 mappers y reducers en Docker Compose.
- Ejemplo de salida en reducer.py:

```
pedro 5 //número de accesos fuera de horario laboral
```

Evidencias Creación de los archivos de configuración



Archivos Mapper.py y Reducer.py

Mapper.py

```
mapper.py > ...
1  #!/usr/bin/env python3
2  import sys
3  import os
4  from datetime import datetime
5
6  # Preparar variables
7  fuera_conteo = 0
8  dentro_conteo = 0
9  fuera_lineas = []
10 dentro_lineas = []
11
12 # Leer y procesar líneas
13 for line in sys.stdin:
14     try:
15         line = line.strip()
16         if not line:
17             continue
18
19         partes = line.split()
20         if len(partes) != 3:
21             continue
22
23         fecha, hora, usuario_raw = partes
24         hora_dt = datetime.strptime(hora, "%H:%M:%S").time()
25         usuario = usuario_raw.split(":")[1] if usuario_raw.startswith("usuario:") else "desconocido"
26
27         if hora_dt < datetime.strptime("08:00:00", "%H:%M:%S").time() or hora_dt > datetime.strptime("18:00:00", "%H:%M:%S").time():
28             fuera_lineas.append(f"{fecha} {hora} usuario:{usuario}")
29             print(f"{usuario}\t1")
30             fuera_conteo += 1
31         else:
32             dentro_lineas.append(f"{fecha} {hora} usuario:{usuario}")
33             dentro_conteo += 1
34     except:
35         continue
36
37 # Crear carpetas
38 os.makedirs("output_usuarios", exist_ok=True)
39
40 # Escribir archivos con encabezado
41 with open("output_usuarios/usuarios_fuera_horario.txt", "w", encoding="utf-8") as f:
42     f.write(f" ♦ Usuarios FUERA del horario laboral: {fuera_conteo} accesos\n")
43     f.write("\n".join(fuera_lineas))
44
45 with open("output_usuarios/usuarios_dentro_horario.txt", "w", encoding="utf-8") as f:
46     f.write(f" ♦ Usuarios DENTRO del horario laboral: {dentro_conteo} accesos\n")
47     f.write("\n".join(dentro_lineas))
48
```

Reducer.py

```
reducer.py > ...
15 total_accesos = sum(conteo.values())
16
17 os.makedirs("output_mapreduce", exist_ok=True)
18
19 with open("output_mapreduce/salida.out", "w", encoding="utf-8") as out:
20     out.write(f" ♦ Consolidado final de accesos fuera de horario por usuario: {total_accesos}")
21     for user, total in conteo.items():
22         out.write(f"{user} {total}\n")
23
```

docker-compose.yml

```
docker-compose.yml
1  services:
2      # Run Service
3      mapper1:
4          build: .
5          volumes:
6              - ../app
7          command: >
8              bash -c "
9                  mkdir -p output_usuarios output_mapreduce &&
10                 cat splits/split_aa | python3 mapper.py > output_mapreduce/map1.txt &&
11                 echo ' * Resultado del map1.txt:' $(wc -l < output_mapreduce/map1.txt) 'registros fuera de horario' > output_mapreduce/map1.out &&
12                 cat output_mapreduce/map1.txt >> output_mapreduce/map1.out
13                 "
14      # Run Service
15      mapper2:
16          build: .
17          volumes:
18              - ../app
19          command: >
20              bash -c "
21                  mkdir -p output_usuarios output_mapreduce &&
22                 cat splits/split_ab | python3 mapper.py > output_mapreduce/map2.txt &&
23                 echo ' * Resultado del map2.txt:' $(wc -l < output_mapreduce/map2.txt) 'registros fuera de horario' > output_mapreduce/map2.out &&
24                 cat output_mapreduce/map2.txt >> output_mapreduce/map2.out
25                 "
26      # Run Service
27      mapper3:
28          build: .
29          volumes:
30              - ../app
31          command: >
32              bash -c "
33                  mkdir -p output_usuarios output_mapreduce &&
34                 cat splits/split_ac | python3 mapper.py > output_mapreduce/map3.txt &&
35                 echo ' * Resultado del map3.txt:' $(wc -l < output_mapreduce/map3.txt) 'registros fuera de horario' > output_mapreduce/map3.out &&
36                 cat output_mapreduce/map3.txt >> output_mapreduce/map3.out
37                 "
38      # Run Service
39      mapper4:
40          build: .
41          volumes:
42              - ../app
43          command: >
44              bash -c "
45                  mkdir -p output_usuarios output_mapreduce &&
46                 cat splits/split_ad | python3 mapper.py > output_mapreduce/map4.txt &&
47                 echo ' * Resultado del map4.txt:' $(wc -l < output_mapreduce/map4.txt) 'registros fuera de horario' > output_mapreduce/map4.out &&
48                 cat output_mapreduce/map4.txt >> output_mapreduce/map4.out
49                 "
50      # Run Service
51      reducer:
52          build: .
53          volumes:
54              - ../app
55          depends_on:
56              - mapper1
57              - mapper2
58              - mapper3
59              - mapper4
60          command: >
61              bash -c "
62                  mkdir -p output_mapreduce &&
63                 cat output_mapreduce/map*.txt | sort | python3 reducer.py
64                 "
```

EJECUCION

Evidencias archivos .txt creados

```
└─ output_mapreduce
   ├── map1.out
   ├── map1.txt
   ├── map2.out
   ├── map2.txt
   ├── map3.out
   ├── map3.txt
   ├── map4.out
   ├── map4.txt
   └── salida.out
└─ output_usuarios
   ├── usuarios_dentro_horario.txt
   └── usuarios_fuera_horario.txt
└─ splits
   ├── split_aa
   ├── split_ab
   ├── split_ac
   └── split_ad
```

Evidencias archivos .out creados

```
└─ output_mapreduce
   ├── map1.out
   ├── map1.txt
   ├── map2.out
   ├── map2.txt
   ├── map3.out
   ├── map3.txt
   ├── map4.out
   ├── map4.txt
   └── salida.out
└─ output_usuarios
   ├── usuarios_dentro_horario.txt
   └── usuarios_fuera_horario.txt
└─ splits
   ├── split_aa
   ├── split_ab
   ├── split_ac
   └── split_ad
```

Salida del consolidado

```
output_mapreduce > ≡ salida.out
1  ♦ Consolidado final de accesos fuera de horario por usuario: 41 accesos
2  ana 9
3  carla 8
4  daniel 8
5  luis 7
6  pedro 9
7  |
```

```
output_usuarios > ≡ usuarios_dentro_horario.txt
1  ♦ Usuarios DENTRO del horario laboral: 14 accesos
2  2025-07-24 08:05:00 usuario:pedro
3  2025-07-24 09:30:00 usuario:luis
4  2025-07-24 11:00:00 usuario:carla
5  2025-07-24 13:00:00 usuario:daniel
6  2025-07-24 15:15:00 usuario:ana
7  2025-07-24 16:50:00 usuario:pedro
8  2025-07-24 17:59:59 usuario:luis
9  2025-07-25 08:00:10 usuario:daniel
10 2025-07-25 09:45:00 usuario:ana
11 2025-07-25 11:30:00 usuario:pedro
12 2025-07-25 13:15:00 usuario:luis
13 2025-07-25 14:50:00 usuario:carla
14 2025-07-25 16:45:00 usuario:daniel
15 2025-07-25 17:59:59 usuario:ana
```

```
output_usuarios > ≡ usuarios_fuera_horario.txt
1  ♦ Usuarios FUERA del horario laboral: 8 accesos
2  2025-07-24 05:30:00 usuario:daniel
3  2025-07-24 07:55:00 usuario:ana
4  2025-07-24 18:01:00 usuario:carla
5  2025-07-24 20:15:00 usuario:daniel
6  2025-07-24 22:30:00 usuario:ana
7  2025-07-25 02:20:00 usuario:pedro
8  2025-07-25 03:45:00 usuario:luis
9  2025-07-25 06:30:00 usuario:carla
```

REPOSITORIO

Colocar el link de Github donde se muestren los *commits* respectivos

https://github.com/josdank/Taller_mapreduce.git

PRESENTACIÓN

Al finalizar la tarea el estudiante deberá subir:

- Un archivo en formato *Nombre_Apellido_Taller4_2B.pdf*.