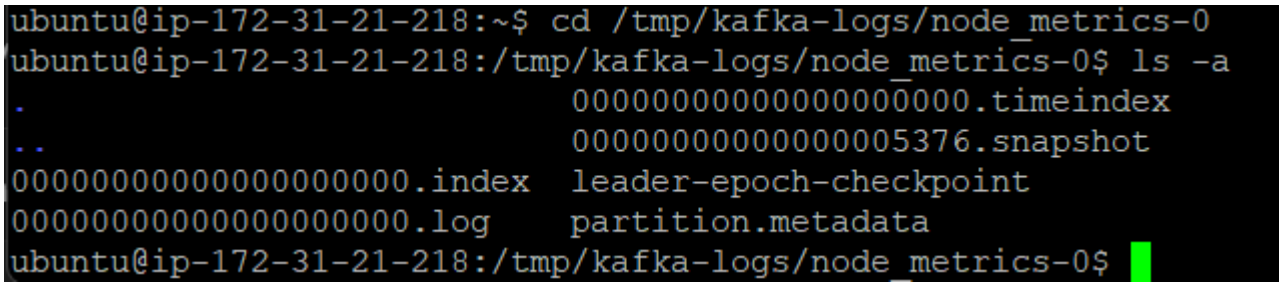
WEEK 4: Finding out the directory where the messages are getting stored and find out how they are being assimilating over time.

- Kafka messages are stored inside "/tmp/kafka-logs"

How to find the location:

cd /tmp/kafka-logs/node_metrics-0

Here, **node_metrics** is the topic name & 0 indicates it's first partition.



What do these files mean?

File Name	Purpose
*.log	The actual data (messages)
*.index	Offset index (maps offsets to physical positions in the log file)
*.timeindex	Timestamp index (maps timestamps to offsets)

Each *.log is a **segment** — Kafka creates new ones as data grows or time/size limits are reached.

Python Script using flask to monitor the messages over time and ship the metrics in a json format to a endpoint:

```
import os
import time
from flask import Flask, jsonify
from threading import Thread

# --- Config ---
LOG_DIR = "/tmp/kafka-logs/node_metrics-0"  # Update this path
CHECK_INTERVAL = 600  # 10 minutes

# Store size history
log_size_history = []

def get_directory_size(directory):
    total = 0
    for dirpath, dirnames, filenames in os.walk(directory):
        for f in filenames:
            fp = os.path.join(dirpath, f)
            total += os.path.getsize(fp)
    return total

def monitor_kafka_logs():
    while True:
        size_bytes = get_directory_size(LOG_DIR)
        timestamp = int(time.time())
        log_size_history.append({
            "timestamp": timestamp,
            "size_bytes": size_bytes
        })
```

```
            print(f"[{time.strftime('%Y-%m-%d %H:%M:%S')}] Size:
{size_bytes / 1024:.2f} KB")
            time.sleep(CHECK_INTERVAL)

    # --- API Server (for Grafana etc.) ---
    app = Flask(__name__)

    @app.route("/metrics")
    def get_metrics():
        return jsonify(log_size_history)

    def run_flask():
        app.run(host="0.0.0.0", port=5000)

    # --- Main ---
    if __name__ == "__main__":
        Thread(target=monitor_kafka_logs, daemon=True).start()
        run_flask()
```

Using **nohup** to run it as a background service:

```
nohup python3 kafka_log_monitor.py &
```

Expose the metrics at:

http://<your-EC2-IP>:5000/metrics

Monitoring the growth of size of kafka logs: