DEMO STEPS

PART 1:

Terminal 1

```
# 1) Start infrastructure
cd ~/Documents/datastream/TP1/project-rag
source ~/Documents/datastream/TP1/datastream/bin/activate
sudo docker compose up -d
```

```
# 2) Clean vector DB (Qdrant)
curl -X DELETE "http://localhost:6333/collections/events"
```

```
# 3) Reset Kafka topics
./reset-topics.sh
```

```
# To see topics (optional)
sudo docker exec -it kafka kafka-topics --bootstrap-server localhost:9092 --list
```

```
# 4) Start consumers
```

Terminal 2

```
cd ~/Documents/datastream/TP1/project-rag
source ~/Documents/datastream/TP1/datastream/bin/activate
python3 index-events.py
```

Terminal 3

```
cd ~/Documents/datastream/TP1/project-rag
source ~/Documents/datastream/TP1/datastream/bin/activate
python3 verify-events.py
```

DURING:

```
# 5) Start one ingestion (producer)
```

Terminal 4

```
cd ~/Documents/datastream/TP1/project-rag
source ~/Documents/datastream/TP1/datastream/bin/activate
python3 ingest-usgs.py
```

```
# 6) Check Kafka stream
```

Terminal 5

```
cd ~/Documents/datastream/TP1/project-rag
source ~/Documents/datastream/TP1/datastream/bin/activate
sudo docker exec -it kafka bash
kafka-console-consumer --bootstrap-server localhost:9092 --topic raw-events --from-beginning --max-messages 10
```

Terminal 6

```
cd ~/Documents/datastream/TP1/project-rag
source ~/Documents/datastream/TP1/datastream/bin/activate
sudo docker exec -it kafka bash
```

```
kafka-console-consumer --bootstrap-server localhost:9092 --topic verified-events
--from-beginning --max-messages 10
```

Terminal 7

```
# 7) Export API key for LLM (before querying)
cd ~/Documents/datastream/TP1/project-rag
source ~/Documents/datastream/TP1/datastream/bin/activate
export GROQ_API_KEY="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"
```

```
# 8) Run RAG query
python3 ask.py "recent earthquakes in California" --minutes 60 --topk 5 --llm
python3 ask.py "recent earthquakes" --topk 5 --llm
```

FOR THE LATENCY:

```
/usr/bin/time -f "Latency: %e seconds" python3 ask.py "recent earthquakes in
California" --minutes 60 --topk 5 --llm
```

New terminal:

```
cd ~/Documents/datastream/TP1/project-rag
source ~/Documents/datastream/TP1/datastream/bin/activate
```

PART 2: Testing commit-after-write

1. Watch lag (group = index-events):

```
sudo docker exec -it kafka kafka-consumer-groups --bootstrap-server localhost:9092
--describe --group index-events
```

```
GROUP          TOPIC          PARTITION  CURRENT-OFFSET  LOG-END-OFFSET  LAG
CONSUMER-ID                                       HOST           CLIENT-ID

index-events    verified-events 0         15             15             0
kafka-python-2.2.15-a2a17394-9847-404e-9d8a-c04e81c1369e /172.18.0.1
kafka-python-2
```

2. Stop Qdrant (cause upsert failure):

```
sudo docker stop qdrant
```

To confirm it is stopped:

```
docker ps | grep qdrant
```

3. Generate new messages (so verified-events grows):

```
python3 ingest-eonet.py
```

(verify-events should pass them through into verified-events)

Lag should increase / not drain while Qdrant is down.

4. Run again:

```
sudo docker exec -it kafka kafka-consumer-groups --bootstrap-server localhost:9092
--describe --group index-events
```

5. Restart Qdrant:

```
sudo docker start qdrant
```

6. After a few min run again:

```
sudo docker exec -it kafka kafka-consumer-groups --bootstrap-server localhost:9092
--describe --group index-events
```

```
luciana@luciana-ZenBook-UX434DA-UM433DA:~/Documents/datastream/TP1/project-rag$
sudo docker exec -it kafka kafka-consumer-groups --bootstrap-server localhost:9092
--describe --group index-events

GROUP         TOPIC           PARTITION  CURRENT-OFFSET  LOG-END-OFFSET  LAG
CONSUMER-ID                                      HOST            CLIENT-ID
index-events    verified-events 0          15              15              0
kafka-python-2.2.15-a2a17394-9847-404e-9d8a-c04e81c1369e /172.18.0.1
kafka-python-2.2.15
luciana@luciana-ZenBook-UX434DA-UM433DA:~/Documents/datastream/TP1/project-rag$
sudo docker exec -it kafka kafka-consumer-groups --bootstrap-server localhost:9092
--describe --group index-events
```

(new ingest running)

```
GROUP         TOPIC           PARTITION  CURRENT-OFFSET  LOG-END-OFFSET  LAG
CONSUMER-ID                                      HOST            CLIENT-ID
index-events    verified-events 0          15              279             264
kafka-python-2.2.15-a2a17394-9847-404e-9d8a-c04e81c1369e /172.18.0.1
kafka-python-2.2.15
luciana@luciana-ZenBook-UX434DA-UM433DA:~/Documents/datastream/TP1/project-rag$
```

(a few min after restarting qrant)

```
GROUP         TOPIC           PARTITION  CURRENT-OFFSET  LOG-END-OFFSET  LAG
CONSUMER-ID                                      HOST            CLIENT-ID
index-events    verified-events 0          282             282             0
kafka-python-2.2.15-a2a17394-9847-404e-9d8a-c04e81c1369e /172.18.0.1
kafka-python-2.2.15
```