BigData – Projekat 2

Student: Uroš Pešić

Profesor: Dragan Stojanović

Kafka Producer App

 Pomoćna aplikacija koja čita vrstu po vrstu iz emission_data.csv fajla, na osnovu kojeg se kreira DTO objekat sa odgovarajućim atributima, koji se serijalizuje u JSON i šalje na kafka topic "traffic-data".

```
public static void produceEvent(KafkaProducer<String, DataEntryDTO> producer, String topic, String lineCsy){
          DataEntrvDTO dataObi = toDTOObiect(lineCsv):
          ProducerRecord<String, DataEntryDTO> rec = new ProducerRecord<>(topic, dataObj.getKey(), dataObj);
          producer.send(rec):
          System.out.println("New event. Event key(Vehicle ID) = " + dataObj.getKey());
public static void main(String[] args) {
          if (args.length < 2) {
          System.out.println("Missing input parameters...");
          System.exit(1);
          String inputPath = args[0]:
          String topic = args[1];
          try (KafkaProducer<String, DataEntryDTO> producer = createProducer()) {
                    File f = new File(inputPath + "/emission data.csv");
                    try (FileReader fr = new FileReader(f)) {
                    try(BufferedReader rd = new BufferedReader(fr)){
                              //skipping 1st line
                              rd.readLine();
                              String line = rd.readLine();
                              while(line != null){
                                        produceEvent(producer, topic, line);
                                       line = rd.readLine();
                                        Thread.sleep(1000);
                    producer.close();
```

Atributi klase DataEntryDTO

Iako se ne koriste svi navedeni atributi ostavio sam neke po izboru.

```
public class DataEntryDTO implements Serializable {
  private int timestep;
  @JsonGetter("timestep")
  public int getTimestep(){
    return this. timestep;
  private float vehicleCO:
  @JsonGetter("vehicle CO")
  public float getVehicleCO(){
    return this. vehicleCO;
  private float vehicleCO2:
  @JsonGetter("vehicle CO2")
  public float getVehicleCO20(
     return this. vehicleCO2;
  private float vehicleHC;
  @JsonGetter("vehicle HC")
  public float getVehicleHC(){
     return this. vehicleHC:
  private float vehicleNOx;
  @JsonGetter("vehicle NOx")
  public float getVehicleNOx(){
     return this. vehicleNOx:
```

```
public class DataEntryDTO implements Serializable {
  private float vehiclePMx;
  @JsonGetter("vehicle PMx")
  public float getVehiclePMx(){
     return this. vehiclePMx;
  private float vehicleFuel:
  @JsonGetter("vehicle fuel")
  public float getVehicleFuelOf
     return this. vehicleFuel;
  private String vehicleID:
  @JsonGetter("vehicle id")
  public String getVehicleID(){
     return this. vehicleID;
  private String vehicleLane:
  @JsonGetter("vehicle lane")
  public String getVehicleLane(){
     return this. vehicleLane:
  private float vehicleNoise;
  @JsonGetter("vehicle noise")
  public float getVehicleNoise(){
     return this. vehicleNoise:
```

```
public class DataEntryDTO implements Serializable {
private String vehicleType;
  @JsonGetter("vehicle type")
  public String getVehicleType(){
    return this. vehicleType;
  private boolean waiting;
  @JsonGetter("vehicle waiting")
  public boolean getVehicleWaiting(){
    return this. waiting;
  private float vehicleX:
  @JsonGetter("vehicle x")
  public float getVehicleX(){
    return this. vehicleX;
  private float vehicleY:
  @JsonGetter("vehicle v")
  public float getVehicleY(){
    return this. vehicleY;
```

Spark Structured Streaming

Komande za pokretanje (nalazimo se u folderu projekta – "BigData-Projekat2")
 docker build --rm -f spark-streaming-app ./SparkStreamingApp/
 docker compose -f ./SparkStreaming/docker-compose.yml up -d --scale spark-worker=2

```
uros@uros-A5-X1:~/Master/BigData/Projekti/Projekat2S docker compose -f ./SparkStreaming/docker-compose.vml up -d --scale spark-worker=2
[+] Runnina 9/13
  Network mobility infr
                                             Created
  Volume "sparkstreaming kafka data"
                                            Created
  Volume "sparkstreaming cassandra data1"
                                            Created
  Volume "sparkstreaming zookeeper data"
                                            Created

✓ Container cassandra-1

✓ Container spark

✓ Container sparkstreaming-spark-worker-2 Started

✓ Container sparkstreaming-spark-worker-1 Started

✓ Container zookeeper

✓ Container kafka

✓ Container init-cassandra

✓ Container consumer-app

✓ Container init-kafka

uros@uros-A5-X1:~/Master/BigData/Proiekti/Proiekat2$ docker ps
CONTAINER ID
                                                                                          STATUS
                                                                                                                         PORTS
                                                                                                                                                                                                     NAMES
053a5b31a78b
               consumer-app
                                           "python -u /consumer..."
                                                                                                                                                                                                     consumer-app
                                                                     About a minute ago
                                                                                          Up 9 seconds
727101428c18
               bitnami/kafka:latest
                                           "/opt/bitnami/script...'
                                                                     About a minute ago
                                                                                          Up About a minute
                                                                                                                         9092/tcp, 0.0.0.0:9094->9094/tcp, :::9094->9094/tcp
                                                                                                                                                                                                     kafka
821bdc4c57a8
               cassandra:latest
                                           "docker-entrypoint.s..."
                                                                                          Up About a minute (healthy)
                                                                                                                         7000-7001/tcp, 7199/tcp, 9042/tcp, 9160/tcp
                                                                                                                                                                                                     cassandra-1
                                                                     About a minute ago
3b13ca19574f
              bitnami/spark:latest
                                           "/opt/bitnami/script..."
                                                                                          Up About a minute
                                                                                                                         0.0.0.0:8080->8080/tcp, :::8080->8080/tcp
                                                                                                                                                                                                     spark
                                                                     About a minute ago
7c5de96147a2
              bitnami/spark:latest
                                           "/opt/bitnami/script..."
                                                                     About a minute ago
                                                                                          Up About a minute
                                                                                                                                                                                                     sparkstreaming-spark-worker-1
8b17a9243491
              bitnami/zookeeper:latest
                                           "/opt/bitnami/script..."
                                                                                          Up About a minute
                                                                                                                         2888/tcp, 3888/tcp, 0.0.0.0:2181->2181/tcp, :::2181->2181/tcp, 8080/tcp
                                                                     About a minute ago
                                                                                                                                                                                                     zookeeper
              bitnami/spark:latest
                                           "/opt/bitnami/script..."
                                                                     About a minute ago
                                                                                          Up About a minute
                                                                                                                                                                                                     sparkstreaming-spark-worker-2
uros@uros-A5-X1:~/Master/BigData/Projekti/Projekat2S
```

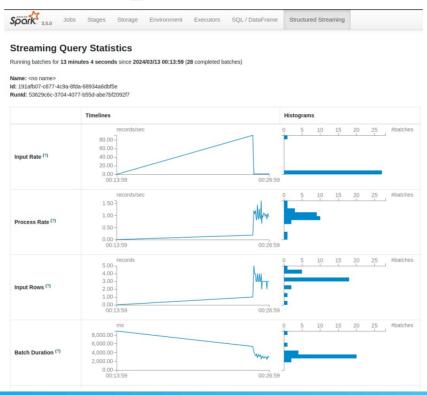
Spark Structured Streaming

```
docker run -d --name spark-streaming-app -p 4040:4040 --network mobility_infr -e SPARK_STREAMING_APP_ARGS_WINDOW_SIZE=1_minute -e SPARK_STREAMING_APP_ARGS_WINDOW_SLIDE=20 seconds spark-streaming-app
```

- Napomena: Argumenti window_size I window_slide moraju da se navedu u formatu
 "duration"_"time_unit", kao na primeru iznad (npr. 1_minute, 5_minutes, 10_seconds, itd.)
 java -jar ./SUMODataProducer/SUMODataProducer.jar ./SUMODataProducer/input traffic data
- Pokrećemo KafkaProducer aplikaciju koja je "zapakovana" kao jar fajl. Potrebno je navesti putanju do input foldera I ime topica na koji se podaci šalju (kako je u obe streaming aplikacija ime topica "traffic data", u data producer aplikaciji ne treba dirati ovaj parametar)

```
kat2$ java -jar ./SUMODataProducer/SUMODataProducer.jar ./SUMODataProd<u>ucer/input traffic data</u>
SLF4J: Failed to load class "org.slf4j.impl.StaticLoggerBinder".
SLF4J: Defaulting to no-operation (NOP) logger implementation
SLF4J: See http://www.slf4j.org/codes.html#StaticLoggerBinder for further details.
New event. Event key(Vehicle ID) = bus0
New event. Event key(Vehicle ID) = motorcycle0
New event. Event key(Vehicle ID) = veh0
New event. Event key(Vehicle ID) = bus0
New event. Event key(Vehicle ID) = motorcycle0
New event. Event key(Vehicle ID) = veh0
New event. Event key(Vehicle ID) = veh1
New event. Event key(Vehicle ID) = bus0
New event. Event key(Vehicle ID) = motorcycle0
New event. Event key(Vehicle ID) = veh0
New event. Event key(Vehicle ID) = veh1
New event. Event key(Vehicle ID) = veh2
New event. Event key(Vehicle ID) = veh3
```

 Napomena: U Spark Streaming aplikaciji je moguće izostaviti argument window_slide i tada ćemo imati tumbling prozore veličine window_size



• Komandom docker run --rm -it --network mobility_infr nuvo/docker-cqlsh cqlsh cassandra-1 9042 -- cqlversion=3.4.6 je moguće dobiti interaktivni shell za izvršavanje upita nad bazom u kojoj su smešteni rezultati obrade. Keyspace se zove *mobility*, ime familije kolona je *traffic_data*.

uros@uros-A5-X1:-/Master/BigData/Projekti/Projekat2\$ docker runrm -itnetwork mobility infr nuvo/docker-cqlsh cqlsh cassandra-1 9042cqlversion=3.4.6													
Connected to Test Cluster at cassandra													
[cqlsh 5.0.1 Cassandra 4.1.3 CQL s	pec 3.4.6 Nati	.ve protocol v5]											
Use HELP for help.													
cqlsh> use mobility;													
cqlsh:mobility> select * from traffic_data;													
10				sum_co2	sum_fuel	sum_hc	sum_nox	sum_pmx		time_to	vehicle_count	zone	
198e3d7a-bf55-43ad-a657-92c44e3f850b	lane_agg	669874507_0	null	null	null	null	null	null	2024-03-13T00:25:40.000Z	2024-03-13T00:26:40.000Z	2	null	
758c4e80-fe99-47d3-a318-e5f6e04ee295	lane agg			null	null		null			2024-03-13T00:26:20.000Z			
aa86d88b-d7db-47d6-9c20-498c5aff43f7			2362.47998	1.4188e+05	45036.80859	46.1	732.54999	20.54				SE	
f8f1bc96-8fea-47b4-b539-cf7735d70f9f	lane_agg			null	null		null			2024-03-13T00:26:40.000Z			
9ea7d83b-1912-4a05-a675-41f95a2bc5c4	lane_agg			null			null		2024-03-13T00:26:00.000Z				
decd2d91-5ac7-4891-8219-3e8f9cb2f6bc	lane agg	260124785#3 1		null	null		null			2024-03-13T00:26:20.000Z			
129ee1b7-ec4f-4b3b-bc18-96d055f69d70	lane_agg			null			null			2024-03-13T00:27:00.000Z			
a59c6724-b8ba-4ce7-83a6-1e242eb95912	lane_agg			null	null		null		2024-03-13T00:25:20.000Z	2024-03-13T00:26:20.000Z	1	null	
9726da26-6ce4-4cc0-a6ec-0f12667c6281	lane_agg			null	null		null		2024-03-13T00:26:00.000Z	2024-03-13T00:27:00.000Z			
ec6d23b9-7907-4e40-a13c-51a298b1fa81	lane_agg	:97839381 1 0		null	null		null		2024-03-13T00:26:00.000Z	2024-03-13T00:27:00.000Z	1	null	
9d2c6a26-587f-447c-90ef-8c05b7f5a8b3	lane_agg	49292938#4 0		null	null		null		2024-03-13T00:25:40.000Z	2024-03-13T00:26:40.000Z	5	null	
bf8f41b4-8865-439c-a74c-807458450012	lane_agg	734712582#2_0		null	null		null		2024-03-13T00:26:00.000Z	2024-03-13T00:27:00.000Z	4	null	
d1a26417-1433-470c-b81a-3b9c517d85b9	lane_agg	741245822#0_0	null	null	null		null		2024-03-13T00:26:00.000Z	2024-03-13T00:27:00.000Z	4	null	
7a8a43f5-6304-4b06-91e4-883a71c4c35c	lane_agg	48651788#0_0		null	null		null		2024-03-13T00:25:00.000Z	2024-03-13T00:26:00.000Z	5	null	
e0f884da-db84-481e-9bcb-2eeed8882d68	pollution_agg		434.20001	48477.49219	15407.83008		218.14		2024-03-13T00:25:20.000Z	2024-03-13T00:26:20.000Z			
29db426e-5a0f-4ebb-9073-6f7110506f76	pollution_agg			23579.46875	7476.08008		184.49001		2024-03-13T00:25:40.000Z	2024-03-13T00:26:40.000Z			
796fea29-5d84-480e-9008-2d2795c6a778	lane_agg	-173493700#2_0		null	null		null		2024-03-13T00:25:00.000Z	2024-03-13T00:26:00.000Z			
2dc6671d-7e95-4b19-be7e-b842cc7350e0	lane_agg	23182833#1_0		null	null		null		2024-03-13T00:25:40.000Z	2024-03-13T00:26:40.000Z			
9fc85500-835f-4777-9d03-7728aaa94e88	lane_agg	50033135#1_0		null			null			2024-03-13T00:26:20.000Z			
7e8882ea-4011-44ba-8cd3-be4742a6ba45	lane_agg	741245822#0_0		null	null		null			2024-03-13T00:26:00.000Z			
68955be8-0f31-4b24-b14c-5d33da222dd0	lane_agg	44463548_0		null	null		null		2024-03-13T00:25:20.000Z	2024-03-13T00:26:20.000Z			
e23e4d18-e475-4d84-b58c-4fb6aec55a5d	lane_agg			null	null		null			2024-03-13T00:26:00.000Z			
1d9f206f-1602-4d87-bc38-074979b24e23	lane_agg			null	null		null			2024-03-13T00:26:20.000Z			
42d75918-449f-4ee2-967d-19f8f797c979	lane_agg			null	null		null			2024-03-13T00:26:00.000Z			
61fa48a4-3718-44fc-a31c-4d37262760b8			2750.19995	1.2289e+05			558.06						
fe147c3b-1090-48a7-ba53-c9de400407fc			672.95001	61658.21094	19546.96094		413.5		2024-03-13T00:26:00.000Z				
5d0e687a-4e97-427a-bfe3-e72fd6e13ed4	lane_agg			null	null		null			2024-03-13T00:26:40.000Z			
38adb59e-36f1-4521-920e-589bd92e40cc	lane_agg			null	null		null			2024-03-13T00:27:00.000Z			
31f137e6-ba5b-4654-b6bd-1279c8b42728				2287.16992	729.51001		0.69						
b2e8bb8e-ce2d-45c8-aeff-aab871875337	lane_agg			null	null		null			2024-03-13T00:27:00.000Z			
bce36a60-1fbe-47fc-ba04-7f71f2571844			633.31	20245.91016	6425.19971		131.7	4.71	2024-03-13T00:24:40.000Z				
0081841b-9b9a-46f0-b0ba-01ddb9ef9444	lane_agg			null	null		null		2024-03-13T00:26:00.000Z				
30161e1f-b9b0-41c0-9fe3-56273548e6bd	pollution_agg		1351.05005	1.0253e+05	32520.16992		590.46997	14.8	2024-03-13T00:26:00.000Z	2024-03-13T00:27:00.000Z			

Flink Streaming

Komande za pokretanje (nalazimo se u folderu projekta – "BigData-Projekat2")
 docker compose -f ./FlinkStreaming/docker-compose.yml up -d --scale taskmanager=2

```
#aster/BigData/Projekti/Projekat2$ docker compose -f ./FlinkStreaming/docker-compose.vml up -d --scale taskmanager=2
uros@uros-A5-X1:~
[+] Running 9/13
  Network mobility infr
                                             Created
  Volume "flinkstreaming zookeeper data"
                                            Created
  Volume "flinkstreaming kafka data"
                                             Created
  Volume "flinkstreaming cassandra data1"
                                            Created

✓ Container cassandra-1

✓ Container jobmanager

✓ Container zookeeper

✓ Container init-cassandra

✓ Container kafka

✓ Container flinkstreaming-taskmanager-2

✓ Container flinkstreaming-taskmanager-1

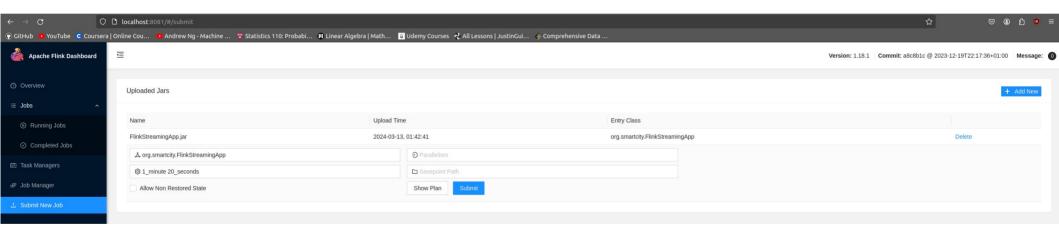
✓ Container consumer-app

✓ Container init-kafka

uros@uros-A5-X1:~/Master/BigData/Projekti/Projekat2$ docker ps
               IMAGE
                                           COMMAND
                                                                     CREATED
                                                                                          STATUS
                                                                                                                         PORTS
                                                                                                                                                                                                      NAMES
CONTAINER ID
350bc7d16669
               consumer-app
                                           "python -u /consumer..."
                                                                                          Up 9 seconds
                                                                                                                                                                                                      consumer-app
                                                                     About a minute ago
3d16e3972aa2
               flink:1.18.1-java8
                                           "/docker-entrypoint...."
                                                                     About a minute ago
                                                                                          Up About a minute
                                                                                                                         6123/tcp, 8081/tcp
                                                                                                                                                                                                      flinkstreaming-taskmanager-1
43a1cd2f69a3
               flink:1.18.1-java8
                                           "/docker-entrypoint..."
                                                                     About a minute ago
                                                                                          Up About a minute
                                                                                                                         6123/tcp, 8081/tcp
                                                                                                                                                                                                      flinkstreaming-taskmanager-2
fc636b0ca02e
               bitnami/kafka:latest
                                            '/opt/bitnami/script..."
                                                                                          Up About a minute
                                                                                                                         9092/tcp, 0.0.0.0:9094->9094/tcp, :::9094->9094/tcp
                                                                                                                                                                                                      kafka
                                                                     About a minute ago
382b47c3daa5
               flink:1.18.1-java8
                                            '/docker-entrypoint..."
                                                                     About a minute ago
                                                                                          Up About a minute
                                                                                                                         6123/tcp, 0.0.0.0:8081->8081/tcp, :::8081->8081/tcp
                                                                                                                                                                                                      jobmanager
                                           "/opt/bitnami/script..."
                                                                                                                         2888/tcp, 3888/tcp, 0.0.0.0:2181->2181/tcp, :::2181->2181/tcp, 8080/tcp
b5a84e1df4dc
               bitnami/zookeeper:latest
                                                                    About a minute ago
                                                                                          Up About a minute
                                                                                                                                                                                                     zookeeper
               cassandra:latest
                                           "docker-entrypoint.s..."
                                                                                          Up About a minute (healthy)
                                                                                                                         7000-7001/tcp, 7199/tcp, 9042/tcp, 9160/tcp
b8a351b14fd9
                                                                    About a minute ago
                                                                                                                                                                                                      cassandra-1
uros@uros-A5-X1:~/Master/BigData/Projekti/Projekat2$
```

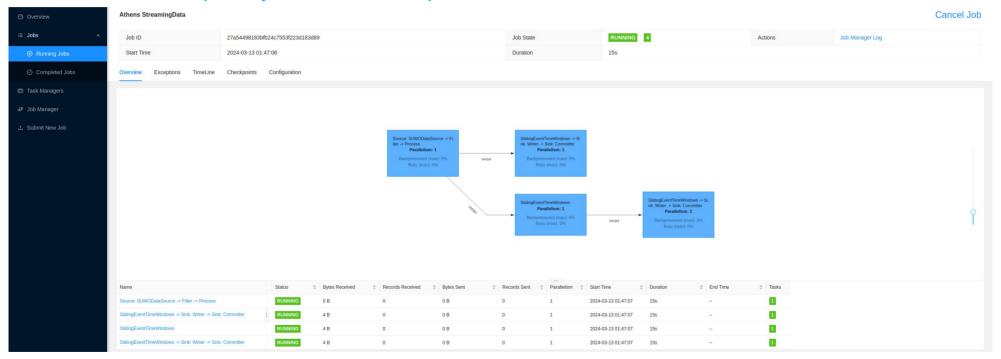
Flink Streaming

 Aplikacija se submit-uje preko web-interfejsa na adresi "localhost:8081". Putanja do jar fajla aplikacije je BigData-Projekat2/FlinkStreaming/FlinkStreamingApp/out/artifacts/FlinkStreamingApp_jar/ FlinkStreamingApp.jar



• Napomena: U Flink streaming aplikaciji je moguće navesti samo jedan argument, čime dobijamo tumbling prozore definisane veličine

 Nakon što pokrenemo streaming aplikaciju, istim komandama od malopre možemo pokrenuti DataProducer aplikaciju, a zatim vršiti upite nad bazom.



.d		'							time_from		vehicle_count	
ea894bf-d2a9-46b8-bc47-6a04c140e47c			748.75		19615.33008		377.99002		2024 03 13 00:49:20			Nh
:6a121d2-5a17-4dff-971e-0b0a20a3691f	lane_agg	44463548_0		null	null	null	null	null	2024 03 13 00:48:20	2024 03 13 00:49:20	4	null
72902d4-78ae-425c-b32c-b455fd4c8fcf	pollution_agg	null	1343.78003	94669.53125	30032.67773	30.57	538.04999	13.96	2024 03 13 00:49:20	2024 03 13 00:50:20	null	SE
f3f2e2e-d3cd-4fc4-a849-8dfad26e3717	pollution_agg	null	34.31	11343.36914	3603.14014	4.99	62.52	2.03	2024 03 13 00:48:40	2024 03 13 00:49:40	null	SW
27053e02-43fc-4ffc-b651-f8fb5baef256	lane_agg	1033862865#0_0			null	null	null	null	2024 03 13 00:49:20	2024 03 13 00:50:20	8	null
2d0d2d2-b184-4680-95d4-324761e28c98	lane_agg	48651788#0_0		null	null	null	null	null	2024 03 13 00:48:40	2024 03 13 00:49:40	6	null
e0541f7-67b5-4cd0-b2ed-ebca65afe7b3	pollution_agg	null	252.53	44383.60156	14078.99023	21.44	292.01001	8.68	2024 03 13 00:49:20	2024 03 13 00:50:20	null	SW
1f5bd9d-9c38-47fd-afb4-cfd247341813	pollution_agg	null	2750.19971	1.2289e+05	39043.85938	44.55	558.06006	18.11	2024 03 13 00:48:40	2024 03 13 00:49:40	null	SE
)f4fea91-82e8-4080-8c0f-c4f06abf2292	pollution_agg	null	626.28998	12337.32031	3935.21973	3.13	5.56	0.3	2024 03 13 00:48:20	2024 03 13 00:49:20	null	NE
:e616ea6-f8bf-40c6-8636-7ac5feb60077	pollution_agg	null	345.32001	41786.60156	13273.73926	16.91	215.73999	6.92	2024 03 13 00:48:40	2024 03 13 00:49:40	null	NW
74bf91b-c5d6-4018-a980-7bcbb5b38015			2548.10986	90073.41406	28644.77148	32.06	332.44	12.39	2024 03 13 00:48:20	2024 03 13 00:49:20	null	SE
2657064-a9b6-44aa-8da5-c2f1c2de4a92	pollution_agg	null	625.88	63313.46094	20108.03906	24.04	326.94	9.91	2024 03 13 00:49:00	2024 03 13 00:50:00	null	NW
59e0b874-742b-4897-ae0f-14dc23c0d94f	pollution_agg	null	220.69	21342.21094	6792.31006	6.18	66	2.36	2024 03 13 00:48:20	2024 03 13 00:49:20	null	NW
014371c-6eb6-4354-b1c4-fc7c722798f3	pollution_agg	null	1670.42993	75525.75781	24089.40039	9.34	31.85	1.48	2024 03 13 00:49:20	2024 03 13 00:50:20	null	
f7027e6-d9c4-4bd4-8d4b-364e8adb9b3b	pollution_agg	null	2962.88965	1.4319e+05	45471.58984	51.82	712.26007	21.79	2024 03 13 00:49:00	2024 03 13 00:50:00		SE
f28c785-372d-4571-b88f-b75a7ac27f93	pollution_agg	null	1889.73999	59872.22656	19096.8418	9.95	26.04	1.26	2024 03 13 00:49:00	2024 03 13 00:50:00	null	
3e7bb2e-2d42-4430-9d23-f70a2f8e54a6	pollution_agg	null	1215.51001	30704.67969	9793.66992	6.25	13.67	0.7	2024 03 13 00:48:40	2024 03 13 00:49:40		NE
e155b15-ee12-4c85-9aba-30c443c4bd62	pollution_agg	null		18293.35938	5804.97021	9.9	123.74001	4.04	2024 03 13 00:49:00	2024 03 13 00:50:00	null	SW
431a6ed-5de4-46d7-9a25-90e8f4f520b6	lane_agg	48651788#0_0		null	null	null	null	null	2024 03 13 00:49:00	2024 03 13 00:50:00	7	null