MSC Data Science

Applied Data Mining

**Preliminary data analysis - Automata-based CER with Flink -** **Logic-based CER with RTEC**

Alevizopoulou Sofia 2022201704002

Avgeros Giannis 2022201704003

Tsiatsios George 2022201704024

Athens 2018

Contents

[Table of figures 5](#_Toc534704652)

[Table of tables 7](#_Toc534704653)

[1. Entity Relation model 8](#_Toc534704654)

[Create tables 9](#_Toc534704655)

[2. Preliminary data analytics 9](#_Toc534704656)

[2.1. Null values 9](#_Toc534704657)

[2.2. SQL commands: 9](#_Toc534704658)

[1. How many ships there are based on shipname? 9](#_Toc534704659)

[2. How many ships have imo number? 9](#_Toc534704660)

[3. How many ships there are based on mmsi( dynamic table)? 10](#_Toc534704661)

[4. How many ships there are based on mmsi (static table)? 10](#_Toc534704662)

[5. How many ships have mouthermmsi (ships that other ships use them as start point)? 10](#_Toc534704663)

[6. Display 5 ships that have mouthermmsi? 11](#_Toc534704664)

[7. Display the most popular mothershipmmsi ship. 11](#_Toc534704665)

[8. Display the ships that have as mothershipmmsi the most popular mothershipmmsi ship 12](#_Toc534704666)

[9. How many ship types there are at the static table? 12](#_Toc534704667)

[10. Display the number of ships based on ship type 13](#_Toc534704668)

[11. Find the shipnames that have no ship type defined 14](#_Toc534704669)

[12. Display the number of vessels per country. The first 3 digits of MMSI declare the region country of the ship. 14](#_Toc534704670)

[13. Display the number of ships and their type per country 15](#_Toc534704671)

[14. Find the ship types for French vessels. 16](#_Toc534704672)

[15. Display the most popular destinations 17](#_Toc534704673)

[16. Display the type of messages that are sent from ship with shiptype=0. 17](#_Toc534704674)

[17. Display the number of messages that are sent from ships based on their type. 18](#_Toc534704675)

[18. Display the number of fishing messages per month 19](#_Toc534704676)

[19. Display the number of messages per month 20](#_Toc534704677)

[20. Display the types of messages inside a fishing area 21](#_Toc534704678)

[21. Display the number of fishing messages inside a fishing area 22](#_Toc534704679)

[22. Display the number of ships that have navigational status “fishing” but out of fishing areas. 23](#_Toc534704680)

[23. Display the number of ships per type that send fishing messages 24](#_Toc534704681)

[24. Display the traffic per month 25](#_Toc534704682)

[25. Display the average speed per Month 26](#_Toc534704683)

[26. Display average speed, draught, length and width per ship type 27](#_Toc534704684)

[27. Display the traffic of Brest port 28](#_Toc534704685)

[28. Display the average of draught, width per month for each ship type 29](#_Toc534704686)

[29. Display the usage of SAR vessels every month 30](#_Toc534704687)

[30. Display the type of atons 31](#_Toc534704688)

[31. Make a plot for a specific ship according its ais messages (python) 32](#_Toc534704689)

[32. Make a plot for a specific ship according its fishing ais messages. All these messages are inside a fishing area (QJIS) 33](#_Toc534704690)

[33. Make a plot for a specific ship according its fishing ais messages. All these messages are inside a fishing area. Fishing areas are presented. (QJIS) 34](#_Toc534704691)

[34. Make a plot for a specific ship according its fishing ais messages. All these messages are inside a fishing area. Fishing areas and constraint fishing areas are presented. (QJIS) 35](#_Toc534704692)

[35. Plot the route of a specific vessel inside a fishing area (python) 36](#_Toc534704693)

[36. Plot constrain fishing areas (python) 37](#_Toc534704694)

[37. Plot fishing areas(python) 38](#_Toc534704695)

[38. Plot ports of Brittany(python) 39](#_Toc534704696)

[39. Plot the ports of the Europe (python) 40](#_Toc534704697)

[2.3. Repository 40](#_Toc534704698)

[3. Automata-based CER with Flink 42](#_Toc534704699)

[3.1. Complex event detecting with automata-based framework 42](#_Toc534704700)

[3.1.1. System Architecture 42](#_Toc534704701)

[3.1.2. System Deployment 42](#_Toc534704702)

[3.1.3. Online Noise Reduction 43](#_Toc534704703)

[3.1.4. Grid partitioning 44](#_Toc534704704)

[3.2. Trajectory Events 45](#_Toc534704705)

[1. Gaps in communication 46](#_Toc534704706)

[2. Vessel speed according its type 47](#_Toc534704707)

[3. Adrift - Course of ground (COG) differentiates from the heading of a vessel 48](#_Toc534704708)

[4. High speed near port 49](#_Toc534704709)

[5. Long Term Stop 50](#_Toc534704710)

[6. Vessels with false status 51](#_Toc534704711)

[3.3. Complex event 52](#_Toc534704712)

[1. Two vessels co-travelling 52](#_Toc534704713)

[2. Fishing Activity 55](#_Toc534704714)

[3. Vessel Rendezvous 57](#_Toc534704715)

[4. Adrift 60](#_Toc534704716)

[5. Package Picking 63](#_Toc534704717)

[6. Loitering 64](#_Toc534704718)

[3.4. Producer of AIS messages for the kafka topic 66](#_Toc534704719)

[3.5. Watermark Pattern 66](#_Toc534704720)

[3.6. Running Apache Flink Jobs 66](#_Toc534704721)

[3.7. Outcomes of the running jobs 69](#_Toc534704722)

[3.8. Empirical Evaluation 70](#_Toc534704723)

[Rendezvous 70](#_Toc534704724)

[Illegal Fishing 70](#_Toc534704725)

[Speed near port 70](#_Toc534704726)

[3.9. Visualization of the detected events 71](#_Toc534704727)

[1. Cotraveling activity 71](#_Toc534704728)

[2. Loitering activity 72](#_Toc534704729)

[3. Adrift activity 73](#_Toc534704730)

[4. High speed near port 74](#_Toc534704731)

[5. Fishing activity 75](#_Toc534704732)

[6. Vessels rendezvous 78](#_Toc534704733)

[3.10. Time of execution 82](#_Toc534704734)

[3.11. Running commands 83](#_Toc534704735)

[3.12. Repository 84](#_Toc534704736)

[3.13. Source code of flink project 84](#_Toc534704737)

[3.14. Running environment 85](#_Toc534704738)

[3.14.1. Computer specifications 85](#_Toc534704739)

[4. Logic-based CER with RTEC 87](#_Toc534704740)

[4.1. Complex event detecting with RTEC 87](#_Toc534704741)

[4.1.1. System Architecture 87](#_Toc534704742)

[4.2. Trajectory events 87](#_Toc534704743)

[4.3. Complex event 88](#_Toc534704744)

[1. Stopped & lowspeed 89](#_Toc534704745)

[2. WithinArea 89](#_Toc534704746)

[3. Underway & anchored or moored 89](#_Toc534704747)

[4. ChangingSpeed 89](#_Toc534704748)

[5. HighSpeedNearCoast 90](#_Toc534704749)

[6. Gap in communication 90](#_Toc534704750)

[7. Adrift 90](#_Toc534704751)

[8. Rendezvous , Fishing 90](#_Toc534704752)

[4.4. Outcomes of the system 90](#_Toc534704753)

[4.4.1. Running patterns for different window size 90](#_Toc534704754)

[4.4.2. Ground truth for detected events with window size: 600 secs 91](#_Toc534704755)

[4.5. Visualization of the detected events 92](#_Toc534704756)

[1. High Speed Near Coast 92](#_Toc534704757)

[2. rendezVous 93](#_Toc534704758)

[3. Adrift 94](#_Toc534704759)

[4.6. Time of execution 95](#_Toc534704760)

[4.7. Running commands 96](#_Toc534704761)

[4.8. Repository 96](#_Toc534704762)

[4.9. Running environment 96](#_Toc534704763)

[4.9.1. Computer specifications 96](#_Toc534704764)

[References 97](#_Toc534704765)

# Table of figures

[Figure 1:Number of mmsis at 2 tables (static, dynamic) 10](#_Toc534704766)

[Figure 2: Ships with mothermmsi 11](#_Toc534704767)

[Figure 3:Ships with the same mothermmmsi 12](#_Toc534704768)

[Figure 4: Ships per type 13](#_Toc534704769)

[Figure 5 Visualization of number of each vessel type. 13](#_Toc534704770)

[Figure 6:Ships with no type defined 14](#_Toc534704771)

[Figure 7:Ships per counrty 14](#_Toc534704772)

[Figure 8:Detailed info for ships per counrty 15](#_Toc534704773)

[Figure 9:Ships of France 16](#_Toc534704774)

[Figure 10:Most popular destinations 17](#_Toc534704775)

[Figure 11:Messages sent from ships with type=0 17](#_Toc534704776)

[Figure 12:Number of messages sent based on ship type 18](#_Toc534704777)

[Figure 13:Fishing messages per month 19](#_Toc534704778)

[Figure 14:Visualization of fishing messages per month 19](#_Toc534704779)

[Figure 15;Number of messages per month 20](#_Toc534704780)

[Figure 16:Type of messages inside a fishing area 21](#_Toc534704781)

[Figure 17:Visualization of total and fishing messages inside a fishing area 23](#_Toc534704782)

[Figure 18:Traffic per month 25](#_Toc534704783)

[Figure 19:Visualization of traffic per month 25](#_Toc534704784)

[Figure 20:Average speed of vessels per month 26](#_Toc534704785)

[Figure 21: Visualization of Average speed of vessels per month 26](#_Toc534704786)

[Figure 22:Average speed, draught, length and width per ship type 27](#_Toc534704787)

[Figure 23:Visualization of average speed, draught, length and width per ship type 28](#_Toc534704788)

[Figure 24:Visualization with the vessels that are connected with Brest Port 28](#_Toc534704789)

[Figure 25:Average of draught,width per month for each ship type 29](#_Toc534704790)

[Figure 26:SAR vessels per month 30](#_Toc534704791)

[Figure 27:Visualization of SAR vessels per month 30](#_Toc534704792)

[Figure 28: Visualization of type of atons 31](#_Toc534704793)

[Figure 29:Route of a specific vessel with mmsi= 228931000 32](#_Toc534704794)

[Figure 30: Route of a specific vessel with mmsi= 228931000 with fishing messages 33](#_Toc534704795)

[Figure 31: Route of a specific vessel with mmsi= 228931000 with fishing messages, fishing areas are presented 34](#_Toc534704796)

[Figure 32:Route of a specific vessel with mmsi= 228931000 with fishing messages, fishing areas and constrained fishing areas are presented (fishing: purple, constraint: red) 35](#_Toc534704797)

[Figure 33:Specific route inside a fishing area for vessel with mmsi= 227741610 36](#_Toc534704798)

[Figure 34: Fishing constraint areas 37](#_Toc534704799)

[Figure 35:Fishing areas 38](#_Toc534704800)

[Figure 36:Ports of Brittany 39](#_Toc534704801)

[Figure 37:Ports all over the world 40](#_Toc534704802)

[Figure 38: System deployment 43](#_Toc534704803)

[Figure 39:Two jobs are running: trajectory and complex events 67](#_Toc534704804)

[Figure 40: Patterns for job Id: trajectory event (part a) 67](#_Toc534704805)

[Figure 41:Patterns for job Id: trajectory event (part b) 68](#_Toc534704806)

[Figure 42: Patterns for job Id: complex event 68](#_Toc534704807)

[Figure 43: Cotraveling activity (a) 71](#_Toc534704808)

[Figure 44: Cotraveling activity with zoom 72](#_Toc534704809)

[Figure 45: Loitering activity 72](#_Toc534704810)

[Figure 46: Adrift activity 73](#_Toc534704811)

[Figure 47: High speed near port, grid of 1.2km x 609.4m 74](#_Toc534704812)

[Figure 48: High speed near port, grid of 4.9km x 4.9km (1545 detected events) 74](#_Toc534704813)

[Figure 49: Fishing activity, 300 – 600 seconds gap in communication (blue:300 secs, purle:600 secs) 75](#_Toc534704814)

[Figure 50: Fishing activity, 600 – 900 seconds gap in communication (purple:600 secs, black:900 secs) 76](#_Toc534704815)

[Figure 51: Fishing activity, 900 - 1200 seconds gap in communication (green:1200 secs, black:900 secs) 77](#_Toc534704816)

[Figure 52: Vessels Rendezvous (blue: grid of 4.9km x 4.9km , purple: grid 1.2km x 609.4m) 78](#_Toc534704817)

[Figure 53: Vessels Rendezvous (black: grid of 152 m x 152m , purple: grid 1.2km x 609.4m) 79](#_Toc534704818)

[Figure 54: Vessels Rendezvous (black: grid of 152 m x 152m , purple: grid 1.2km x 609.4m) with zoom 79](#_Toc534704819)

[Figure 55:Vessels Rendezvous (black: grid of 152 m x 152m , yellow: grid 38,2 m x 19m) 80](#_Toc534704820)

[Figure 56:Vessels Rendezvous (black: grid of 152 m x 152m , yellow: grid 38,2 m x 19m) with zoom 81](#_Toc534704821)

[Figure 57:Vessels Rendezvous with zoom for all different grid values 82](#_Toc534704822)

[Figure 58: Cpu info 85](#_Toc534704823)

[Figure 59:Memory info 86](#_Toc534704824)

[Figure 60: High speed near coast (blue: RTEC, red :flink) 92](#_Toc534704825)

[Figure 61: High speed near coast in zoom (blue: RTEC, red :flink) 93](#_Toc534704826)

[Figure 62: Vessels Rendezvous (blue: RTEC, red :flink) 93](#_Toc534704827)

[Figure 63:Vessels Rendezvous in zoom (blue: RTEC, red: flink) 94](#_Toc534704828)

[Figure 64: Adrift from RTEC 94](#_Toc534704829)

[Figure 65:Adrift activity for mmsi 228064900 (blue: RTEC, red: flink) 95](#_Toc534704830)

# Table of tables

[Table 1: Gap Pattern 46](#_Toc534704831)

[Table 2: Too slow or too fast vessel pattern 47](#_Toc534704832)

[Table 3: Heading and CoG pattern 48](#_Toc534704833)

[Table 4:High Speed near port 49](#_Toc534704834)

[Table 5: Long Term Stop 50](#_Toc534704835)

[Table 6: Vessels with false status 51](#_Toc534704836)

[Table 7: Cotravel for 2 vessels pattern – trajectory events 53](#_Toc534704837)

[Table 8: Cotravel for 2 vessels pattern – complex events 54](#_Toc534704838)

[Table 9: Fishing pattern 55](#_Toc534704839)

[Table 10: Vessel Rendezvous pattern - trajectory events 58](#_Toc534704840)

[Table 11: Vessel Rendezvous pattern - complex events 59](#_Toc534704841)

[Table 12: Big difference between heading and course – trajectory event 61](#_Toc534704842)

[Table 13: Problematic route - complex events 62](#_Toc534704843)

[Table 14: Watermark pattern 66](#_Toc534704844)

# Entity Relation model

The tables that we are going to use in this assignment are listed below. They contain real-world data from the maritime domain [https://zenodo.org/record/1167595#.W9BgcFUzapo] which monitor, analyze and visualize the sea movements. There are tables that contain info of the whole maritime activities and their impact on the environment, tables that contain a set of complementary data having spatial and temporal information and tables with information about ships positions within Celtic sea, the Channel and Bay of Biscay (France).

More specifically, there are four categories of data: Navigation data, vessel-oriented data, geographic data, and environmental data. It covers a time span of six months, from October 1st, 2015 to March 31st, 2016.

Tables are:

* aton
* pg\_catalog
* country\_codes
* pg\_temp\_1
* geographic\_features
* pg\_toast.
* geography\_columns
* pg\_toast\_temp\_1
* geometry\_columns
* ports.
* information\_schema
* public.
* raster\_overviews
* receiver
* ship\_types\_detailled\_list
* natura2000
* ship\_types\_list
* navigational\_status
* spatial\_ref\_sys
* raster\_columns
* nari\_ais\_static
* nari\_dynamic
* nari\_dynamic\_aton
* nari\_dynamic\_sar

We are going to use most of the tables at the next section at which we will try to understand the data.

## Create tables

We have added all the above tables at the database.

# 2. Preliminary data analytics

## 2.1. Null values

By observing the data, we discovered many columns that contain multiple null values. We decided it would be a good practice not to remove them in this step, but include them in the dataset and decide in one of the next steps how to handle those cases.

Preliminary data analysis refers to some simple and basic analysis by running some SQL queries. This analysis will helpful in understanding the data.

## 2.2. SQL commands:

### How many ships there are based on shipname?

SELECT *Count* (DISTINCT shipname)   
FROM   nari\_ais\_static;

Count: 4824

Note: Extract this info from static information about ships

### How many ships have imo number?

SELECT *Count* (DISTINCT imo)   
FROM   nari\_ais\_static;

Count: 4033

Note: There are ships with no imo number

### How many ships there are based on mmsi( dynamic table)?

SELECT *Count* (DISTINCT mmsi)   
FROM   nari\_dynamic;

Count: 5055

Note: Extract this info from dynamic information about ships (ais messages). Here we can see that there are ships which send ais messages but there is no info about them at the static table of ships (nari\_static)

### How many ships there are based on mmsi (static table)?

SELECT *Count* (DISTINCT sourcemmsi)   
FROM   nari\_ais\_static;

Count: 4842

Figure 1:Number of mmsis at 2 tables (static, dynamic)

### How many ships have mouthermmsi (ships that other ships use them as start point)?

SELECT *Count* (DISTINCT mothershipmmsi)   
FROM   nari\_ais\_static;

Count: 228

### Display 5 ships that have mouthermmsi?

SELECT shipname,   
       callsign,   
       imo,   
       sourcemmsi,   
       mothershipmmsi,   
       detailed\_type   
FROM   public.nari\_ais\_static,   
       public.ship\_types\_list,   
       public.ship\_types\_detailled\_list   
WHERE  nari\_ais\_static.shiptype = ship\_types\_detailled\_list.id\_detailedtype   
       AND ship\_types\_detailled\_list.id\_shiptype = ship\_types\_list.id\_shiptype   
       AND nari\_ais\_static.mothershipmmsi > 0   
LIMIT  5;

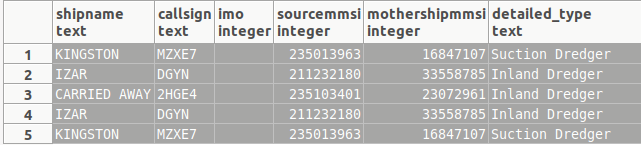


Figure 2: Ships with mothermmsi

### Display the most popular mothershipmmsi ship.

SELECT mothershipmmsi,   
       *Count*(DISTINCT sourcemmsi) AS num   
FROM   nari\_ais\_static   
WHERE  mothershipmmsi > 0   
GROUP  BY mothershipmmsi   
ORDER  BY num DESC;

Mothershipmmsi | count

6320258 | 6

4223106 | 5

### Display the ships that have as mothershipmmsi the most popular mothershipmmsi ship

SELECT DISTINCT sourcemmsi,   
                shipname,   
                callsign,   
                imo,   
                mothershipmmsi,   
                detailed\_type   
FROM   PUBLIC.nari\_ais\_static,   
       PUBLIC.ship\_types\_list,   
       PUBLIC.ship\_types\_detailled\_list   
WHERE  nari\_ais\_static.shiptype = ship\_types\_detailled\_list.id\_detailedtype   
       AND ship\_types\_detailled\_list.id\_shiptype = ship\_types\_list.id\_shiptype   
       AND nari\_ais\_static.mothershipmmsi = 6320258;

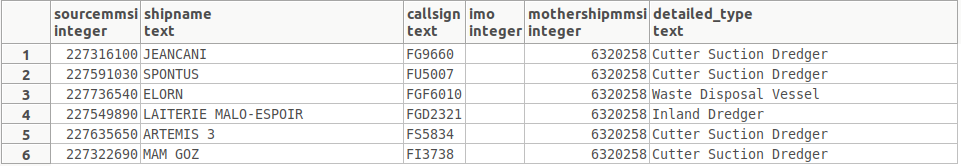


Figure 3:Ships with the same mothermmmsi

### How many ship types there are at the static table?

SELECT *Count* (DISTINCT shiptype)   
FROM   nari\_ais\_static;

Count: 45

### Display the number of ships based on ship type

SELECT type\_name,   
       *Count*(DISTINCT sourcemmsi) AS num   
FROM   PUBLIC.nari\_ais\_static,   
       PUBLIC.ship\_types\_list   
WHERE  ship\_types\_list.shiptype\_min <= nari\_ais\_static.shiptype   
       AND ship\_types\_list.shiptype\_max >= nari\_ais\_static.shiptype   
GROUP  BY type\_name   
ORDER  BY num DESC;

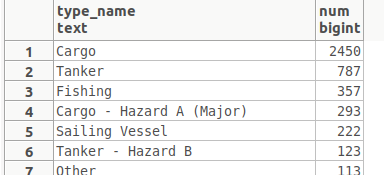


Figure 4: Ships per type

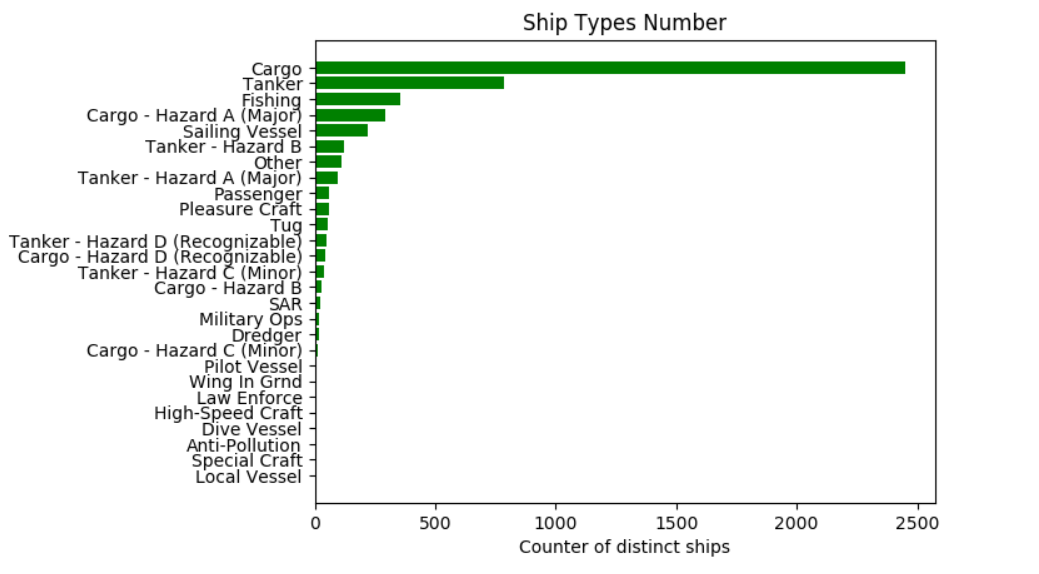


Figure 5 Visualization of number of each vessel type.

### Find the shipnames that have no ship type defined

SELECT DISTINCT shipname   
FROM   nari\_ais\_static   
WHERE  nari\_ais\_static.shiptype IS NULL;

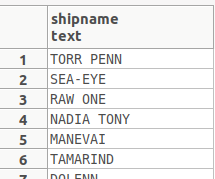


Figure 6:Ships with no type defined

### Display the number of vessels per country. The first 3 digits of MMSI declare the region country of the ship.

SELECT **Count**(DISTINCT sourcemmsi),   
       country   
FROM   public.country\_codes,   
       public.nari\_ais\_static   
WHERE  **Left**(nari\_ais\_static.sourcemmsi :: text, 3) :: INTEGER =   
       country\_codes.mmsi\_country\_code   
GROUP  BY country;

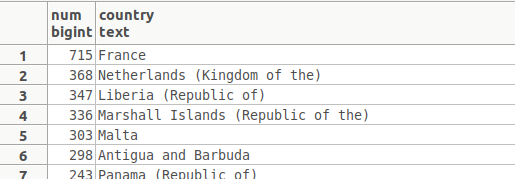


Figure 7:Ships per counrty

### Display the number of ships and their type per country

SELECT **Count**(DISTINCT( sourcemmsi )),   
       country,   
       ais\_type\_summary   
FROM   public.country\_codes,   
       public.nari\_ais\_static,   
       public.ship\_types\_list   
WHERE  **Left**(nari\_ais\_static.sourcemmsi :: text, 3) :: INTEGER =   
              country\_codes.mmsi\_country\_code   
       AND ship\_types\_list.shiptype\_min <= nari\_ais\_static.shiptype   
       AND ship\_types\_list.shiptype\_max >= nari\_ais\_static.shiptype   
GROUP  BY country,   
          ais\_type\_summary;

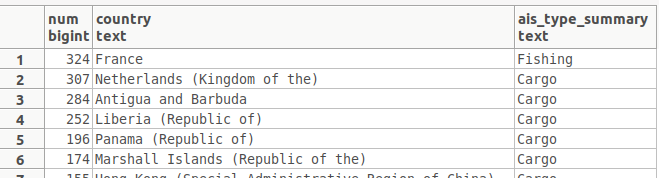


Figure 8:Detailed info for ships per counrty

### Find the ship types for French vessels.

SELECT DISTINCT mmsi\_country\_code   
FROM   PUBLIC.country\_codes   
WHERE  country LIKE 'France';

Note: mmsi codes for France are: 226, 227,228

SELECT shiptype,   
       detailed\_type,   
       **Count**(DISTINCT shipname)   
FROM   public.nari\_ais\_static,   
       public.ship\_types\_list,   
       public.ship\_types\_detailled\_list   
WHERE  nari\_ais\_static.shiptype = ship\_types\_detailled\_list.id\_detailedtype   
       AND ship\_types\_detailled\_list.id\_shiptype = ship\_types\_list.id\_shiptype   
       AND ( **Left**(nari\_ais\_static.sourcemmsi :: text, 3) :: INTEGER = 226   
              OR **Left**(nari\_ais\_static.sourcemmsi :: text, 3) :: INTEGER = 227   
              OR **Left**(nari\_ais\_static.sourcemmsi :: text, 3) :: INTEGER = 228 )   
GROUP  BY shiptype,   
          detailed\_type;

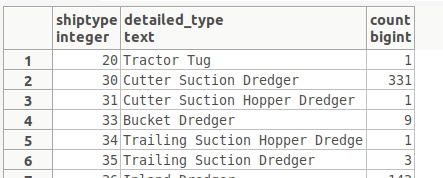


Figure 9:Ships of France

### Display the most popular destinations

SELECT destination,   
       *Count*(DISTINCT sourcemmsi) AS num   
FROM   nari\_ais\_static   
WHERE  **Length**(destination) > 0   
GROUP  BY destination   
ORDER  BY num DESC;

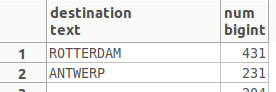


Figure 10:Most popular destinations

### Display the type of messages that are sent from ship with shiptype=0.

SELECT   *Count*(\*),   
         n.status   
FROM     (   
                SELECT mmsi,   
                       status   
                FROM   PUBLIC.nari\_dynamic) a,   
         (   
                         SELECT DISTINCT(sourcemmsi),   
                                         shiptype   
                         FROM            nari\_ais\_static) b,   
         navigational\_status AS n mmsi=sourcemmsi   
AND      b.shiptype=0   
AND      n.code=a.status   
GROUP BY n.status;

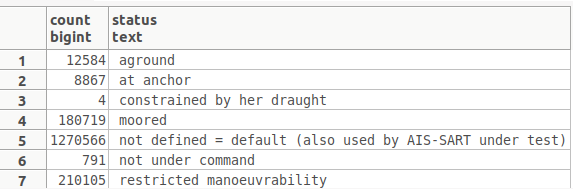


Figure 11:Messages sent from ships with type=0

### Display the number of messages that are sent from ships based on their type.

SELECT *Count*(\*),   
       n.status,   
       shl.ais\_type\_summary   
FROM   (SELECT mmsi,   
               status   
        FROM   PUBLIC.nari\_dynamic) a,   
       (SELECT DISTINCT( sourcemmsi ),   
                       shiptype   
        FROM   nari\_ais\_static) b,   
       ship\_types\_list AS shl,   
       navigational\_status AS n   
WHERE  mmsi = sourcemmsi   
       AND shl.shiptype\_min <= b.shiptype   
       AND shl.shiptype\_max >= b.shiptype   
       AND b.shiptype != 0   
       AND n.code = a.status   
GROUP  BY n.status,   
          shl.ais\_type\_summary;

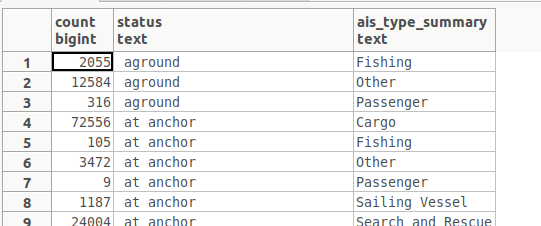


Figure 12:Number of messages sent based on ship type

### Display the number of fishing messages per month

SELECT   *Count*(\*) AS num,   
         m\_month   
FROM     (   
                SELECT status,   
                       date\_part('month',timestamp 'epoch' + t \* interval '1 second') AS m\_month   
                FROM   nari\_dynamic) m   
WHERE    m.status=7   
GROUP BY m\_month   
ORDER BY num DESC;

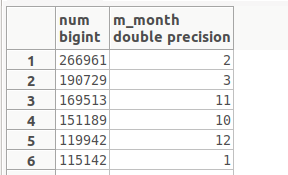


Figure 13:Fishing messages per month

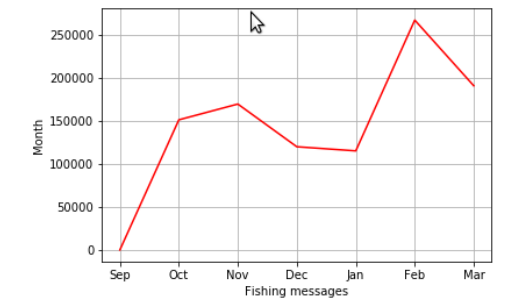


Figure 14:Visualization of fishing messages per month

### Display the number of messages per month

SELECT   *Count*(\*) AS num,   
         m\_month,   
         n.status   
FROM     (   
                SELECT status,   
                       date\_part('month',timestamp 'epoch' + t \* interval '1 second') AS m\_month   
                FROM   nari\_dynamic) m,   
         navigational\_status AS n   
WHERE    m.status=n.code   
GROUP BY m\_month,   
         n.status   
ORDER BY num DESC;

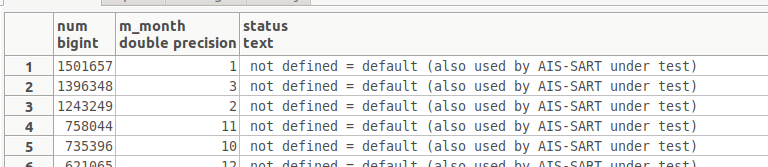


Figure 15;Number of messages per month

### Display the types of messages inside a fishing area

SELECT *Count*(\*) AS num,   
       n.status   
FROM   (SELECT DISTINCT maxlong,   
                        maxlat,   
                        minlong,   
                        minlat   
        FROM   geographic\_features.fishing\_areas\_eu) g,   
       (SELECT lat,   
               lon,   
               status   
        FROM   nari\_dynamic) d,   
       navigational\_status AS n   
WHERE  maxlong >= lon   
       AND minlong <= lon   
       AND maxlat >= lat   
       AND minlat <= lat   
       AND n.code = d.status   
GROUP  BY n.status   
ORDER  BY num DESC;

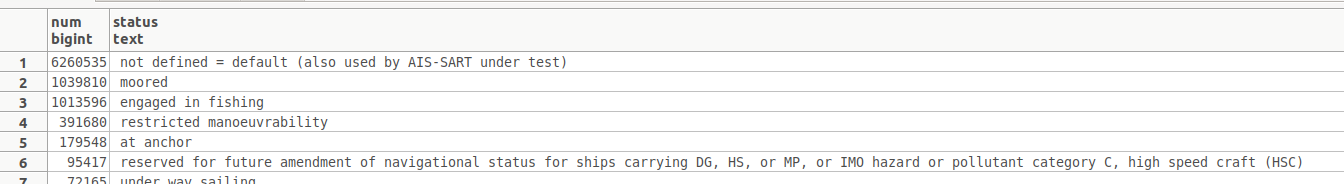


Figure 16:Type of messages inside a fishing area

### Display the number of fishing messages inside a fishing area

FROM   (SELECT DISTINCT maxlong,   
                        maxlat,   
                        minlong,   
                        minlat   
        FROM   geographic\_features.fishing\_areas\_eu) g,   
       (SELECT lat,   
               lon,   
               status   
        FROM   nari\_dynamic) d,   
       navigational\_status AS n   
WHERE  maxlong >= lon   
       AND minlong <= lon   
       AND maxlat >= lat   
       AND minlat <= lat   
       AND n.code = d.status   
       AND n.code = 7;

Count: 1013596

Whereas the total number of messages inside this area is:

SELECT *Count*(\*)   
FROM   (SELECT DISTINCT maxlong,   
                        maxlat,   
                        minlong,   
                        minlat   
        FROM   geographic\_features.fishing\_areas\_eu) g,   
       (SELECT lat,   
               lon,   
               status   
        FROM   nari\_dynamic) d,   
       navigational\_status AS n   
WHERE  maxlong >= lon   
       AND minlong <= lon   
       AND maxlat >= lat   
       AND minlat <= lat   
       AND n.code = d.status;

Count: 9092197

Note: fishing messages is 11,147977% of the total messages

Figure 17:Visualization of total and fishing messages inside a fishing area

### Display the number of ships that have navigational status “fishing” but out of fishing areas.

SELECT *Count*(\*) AS num,   
       n.status   
FROM   (SELECT DISTINCT maxlong,   
                        maxlat,   
                        minlong,   
                        minlat   
        FROM   geographic\_features.fishing\_areas\_eu) g,   
       (SELECT lat,   
               lon,   
               status   
        FROM   nari\_dynamic) d,   
       navigational\_status AS n   
WHERE  ( maxlong < lon   
       OR minlong > lon )   
       AND (maxlat < lat

OR  minlat > lat)   
       AND n.code = d.status

AND n.code = 7  
GROUP  BY n.status   
ORDER  BY num DESC;

Count:0

Note: All ships that have navigational status “fishing” are inside fishing areas.

### Display the number of ships per type that send fishing messages

SELECT type\_name,   
       *Count*(DISTINCT sourcemmsi) AS num   
FROM   PUBLIC.nari\_ais\_static,   
       PUBLIC.ship\_types\_list,   
       PUBLIC.nari\_dynamic   
WHERE  ship\_types\_list.shiptype\_min <= nari\_ais\_static.shiptype   
       AND ship\_types\_list.shiptype\_max >= nari\_ais\_static.shiptype   
       AND status = 7   
GROUP  BY type\_name   
ORDER  BY num DESC;

### Display the traffic per month

SELECT *Count*(mmsi),   
       **Extract**(month FROM **To\_timestamp**(t)) a   
FROM   nari\_dynamic   
GROUP  BY a   
ORDER  BY *Count*(mmsi) DESC;



Figure 18:Traffic per month

Figure 19:Visualization of traffic per month

### Display the average speed per Month

SELECT *Avg*(speed),   
       **Extract**(month FROM **To\_timestamp**(t)) a   
FROM   nari\_dynamic   
GROUP  BY a;

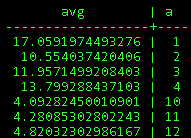


Figure 20:Average speed of vessels per month

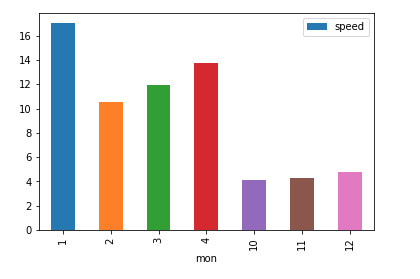


Figure 21: Visualization of Average speed of vessels per month

### Display average speed, draught, length and width per ship type

SELECT b.shiptype,   
       type\_name,   
       *Avg*(speed),   
       *Avg*(b.draught),   
       *Avg*(b.to\_stern),   
       *Avg*(b.to\_starboard)   
FROM   (SELECT mmsi,   
               speed   
        FROM   nari\_dynamic) a,   
       (SELECT DISTINCT( sourcemmsi ),   
                       shiptype,   
                       draught,   
                       to\_stern,   
                       to\_starboard   
        FROM   nari\_ais\_static) b,   
       (SELECT type\_name,   
               shiptype\_min   
        FROM   ship\_types\_list   
        WHERE  shiptype\_min > 0   
               AND shiptype\_max < 100) c   
WHERE  a.mmsi = b.sourcemmsi   
       AND c.shiptype\_min = b.shiptype   
GROUP  BY b.shiptype,   
          type\_name;

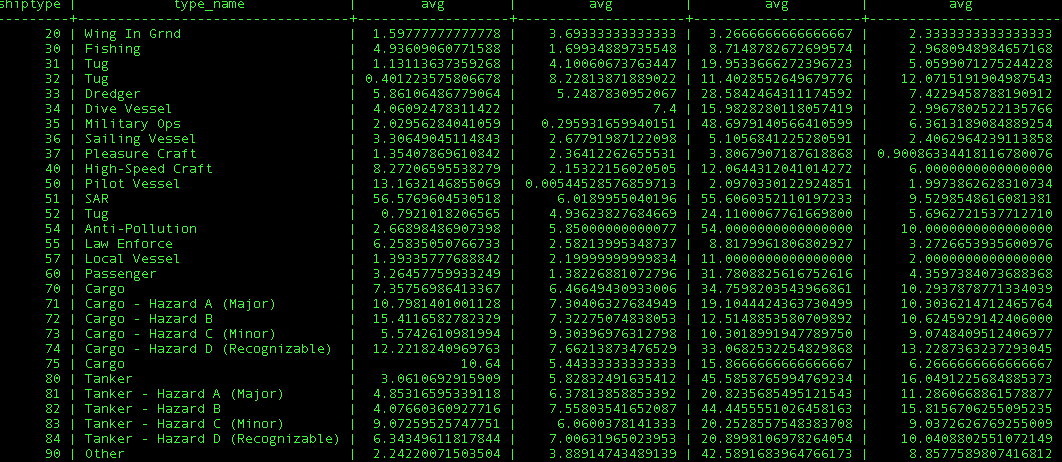
******

Figure 22:Average speed, draught, length and width per ship type

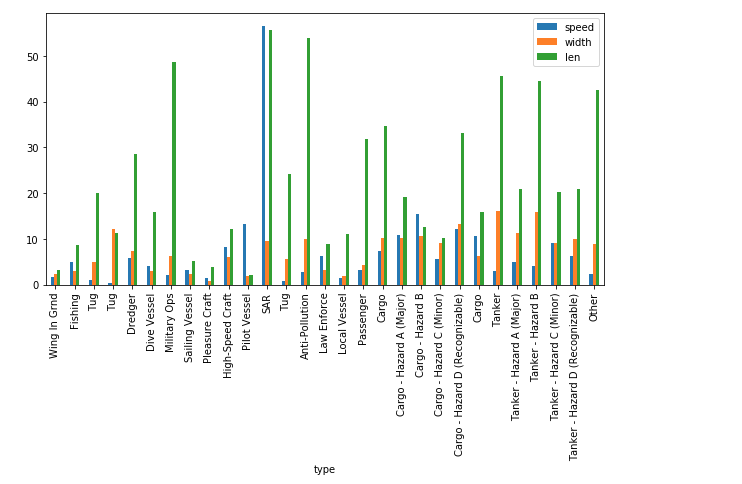
******

Figure 23:Visualization of average speed, draught, length and width per ship type

### Display the traffic of Brest port

SELECT *Count*(\*)   
FROM   nari\_ais\_static   
WHERE  destination LIKE '%BREST%'   
        OR destination LIKE '%BES'   
        OR destination LIKE ' ';

*Count of vessels connected with Brest port :640060*

Count of all vessels; 1078617

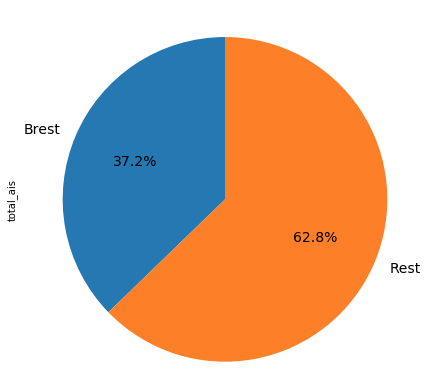


Figure 24:Visualization with the vessels that are connected with Brest Port

### Display the average of draught, width per month for each ship type

SELECT b.shiptype,   
       type\_name,   
       *Avg*(b.draught)      avg\_draught,   
       *Avg*(b.to\_starboard) avg\_len,   
       a.c   
FROM   (SELECT mmsi,   
               speed,   
               **Extract**(month FROM **To\_timestamp**(t)) c   
        FROM   nari\_dynamic) a,   
       (SELECT DISTINCT( sourcemmsi ),   
                       shiptype,   
                       draught,   
                       to\_starboard   
        FROM   nari\_ais\_static) b,   
       (SELECT type\_name,   
               shiptype\_min   
        FROM   ship\_types\_list   
        WHERE  shiptype\_min > 0   
               AND shiptype\_max <= 100) c   
WHERE  a.mmsi = b.sourcemmsi   
       AND c.shiptype\_min = b.shiptype   
GROUP  BY b.shiptype,   
          type\_name,   
          a.c,   
          type\_name   
ORDER  BY a.c;

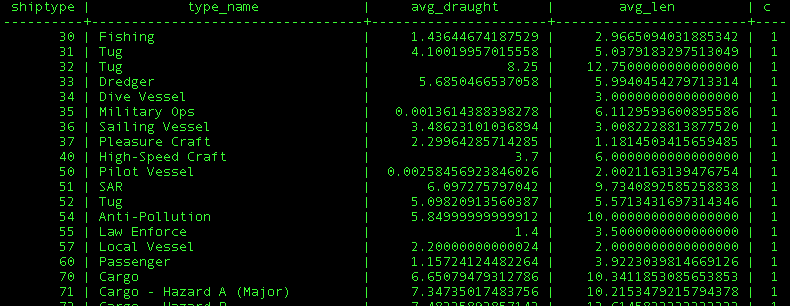
******

Figure 25:Average of draught,width per month for each ship type

### Display the usage of SAR vessels every month

SELECT *Count*(mmsi),   
       **Extract**(month FROM **To\_timestamp**(t)) a,   
       *Avg*(speed)   
FROM   nari\_dynamic\_sar   
GROUP  BY a   
ORDER  BY *Count*(mmsi) DESC;

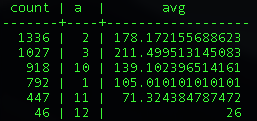
****

Figure 26:SAR vessels per month

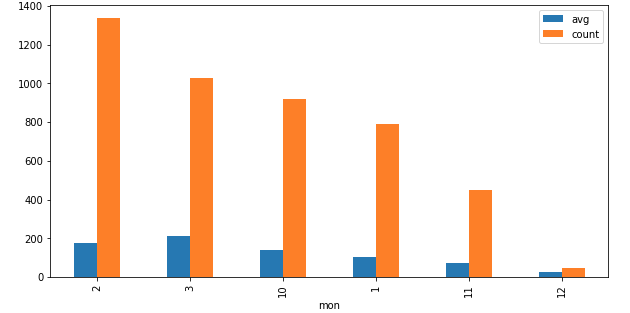
****

Figure 27:Visualization of SAR vessels per month

### Display the type of atons

SELECT typeofaid,   
       *Count*(typeofaid),   
       at.definition,   
       virtual   
FROM   nari\_dynamic\_aton a,   
       aton at   
WHERE  a.typeofaid = at.code   
GROUP  BY typeofaid,   
          at.definition,   
          virtual;

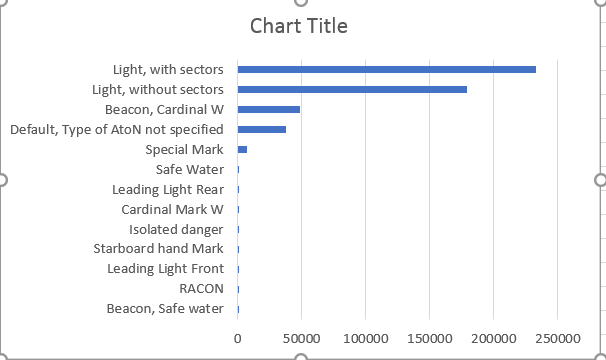


Figure 28: Visualization of type of atons

### Make a plot for a specific ship according its ais messages (python)

Source code: vessel\_route.py

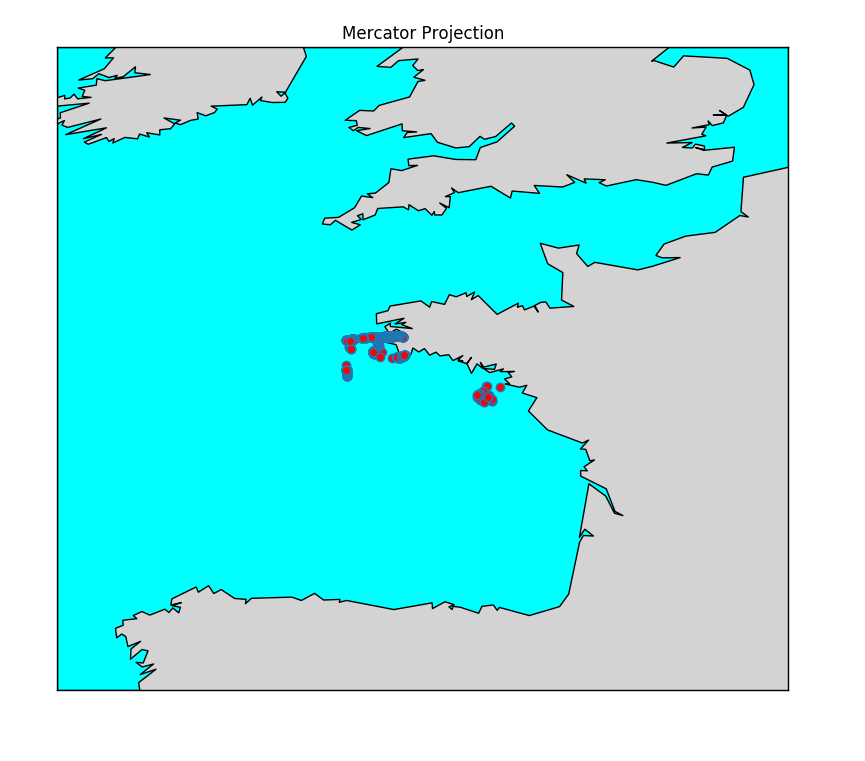


Figure 29:Route of a specific vessel with mmsi= 228931000

### Make a plot for a specific ship according its fishing ais messages. All these messages are inside a fishing area (QJIS)

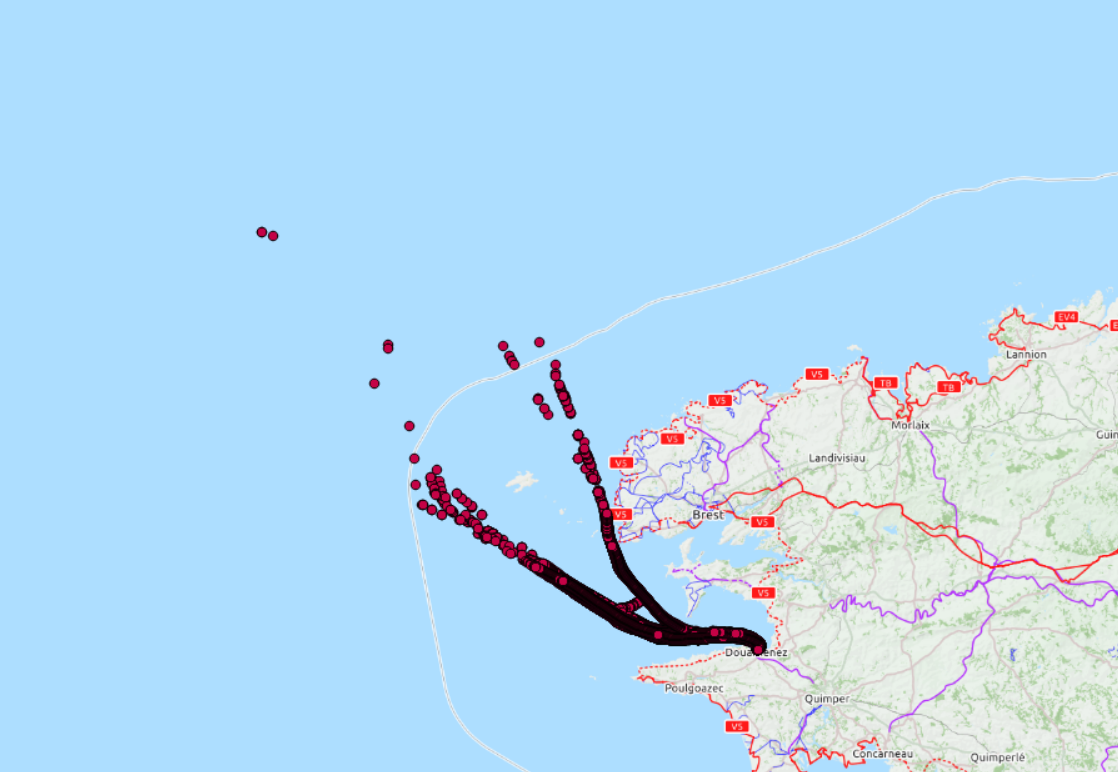


Figure 30: Route of a specific vessel with mmsi= 228931000 with fishing messages

### Make a plot for a specific ship according its fishing ais messages. All these messages are inside a fishing area. Fishing areas are presented. (QJIS)

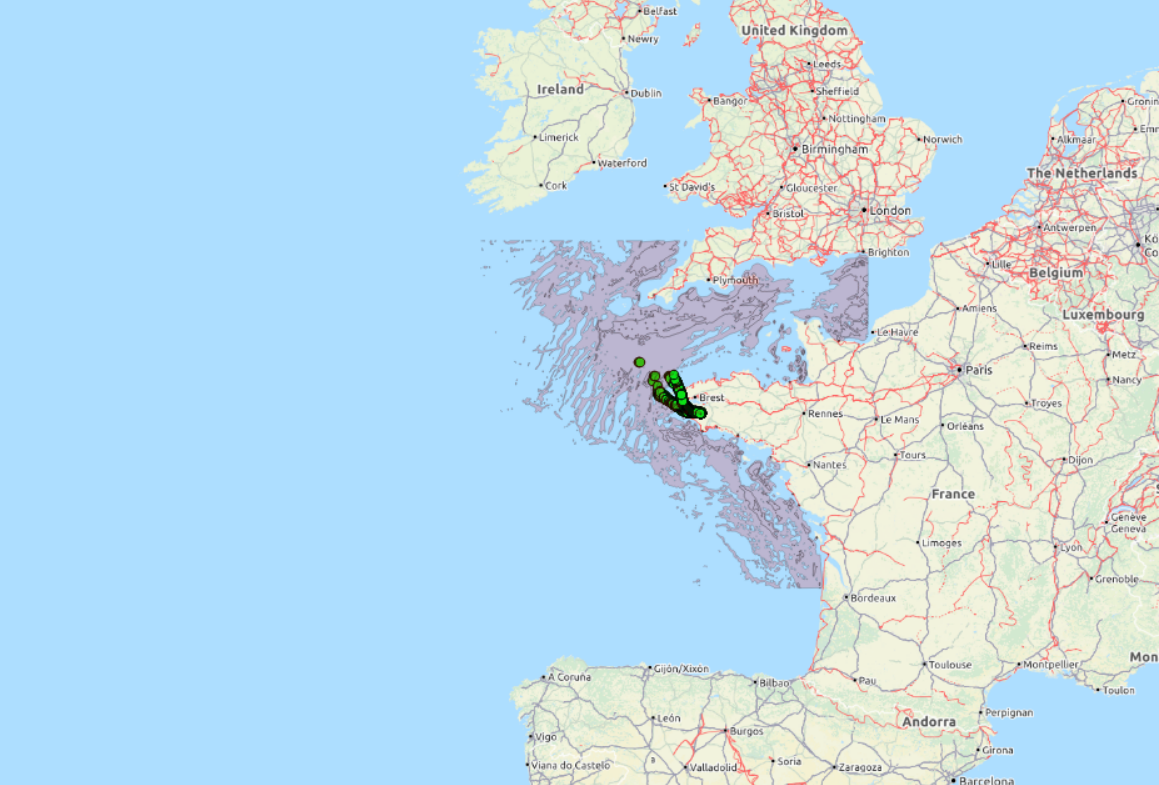


Figure 31: Route of a specific vessel with mmsi= 228931000 with fishing messages, fishing areas are presented

### Make a plot for a specific ship according its fishing ais messages. All these messages are inside a fishing area. Fishing areas and constraint fishing areas are presented. (QJIS)



Figure 32:Route of a specific vessel with mmsi= 228931000 with fishing messages, fishing areas and constrained fishing areas are presented (fishing: purple, constraint: red)

### Plot the route of a specific vessel inside a fishing area (python)

Source code: vessel\_at\_fishing\_area.py

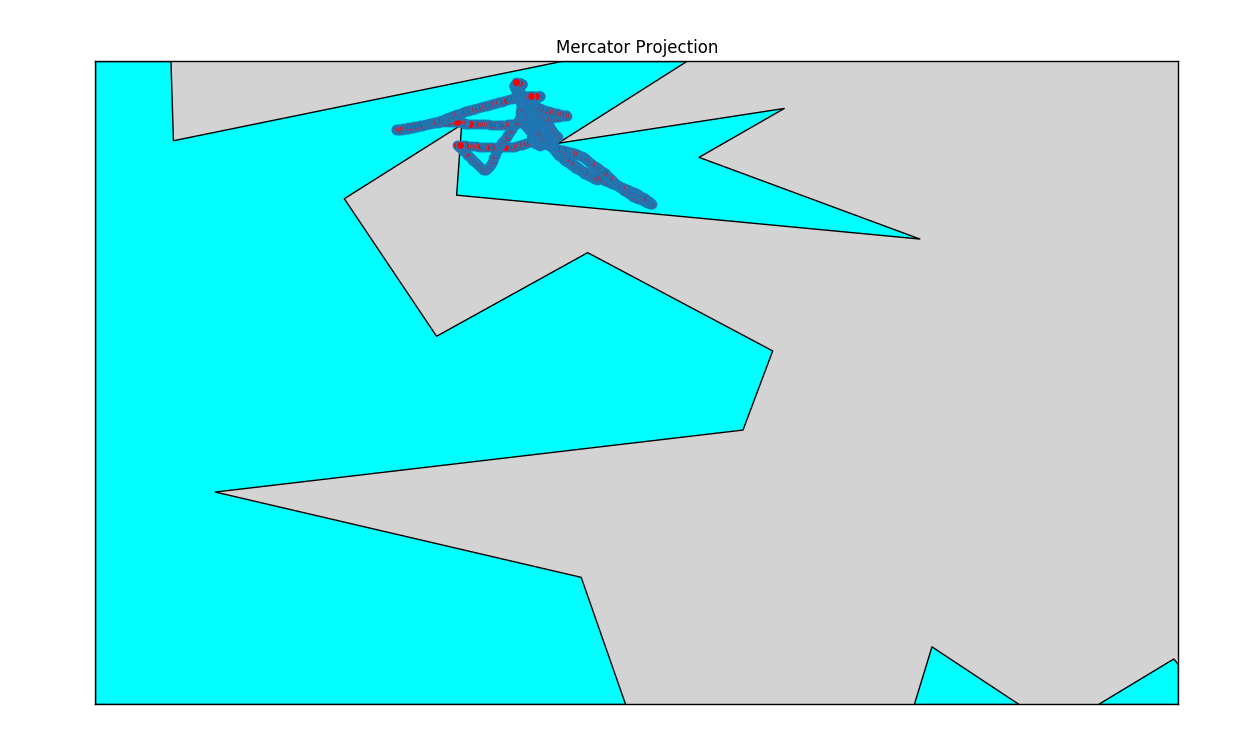


Figure 33:Specific route inside a fishing area for vessel with mmsi= 227741610

### Plot constrain fishing areas (python)

Source code: constrain\_fishing/constrain\_fish.py

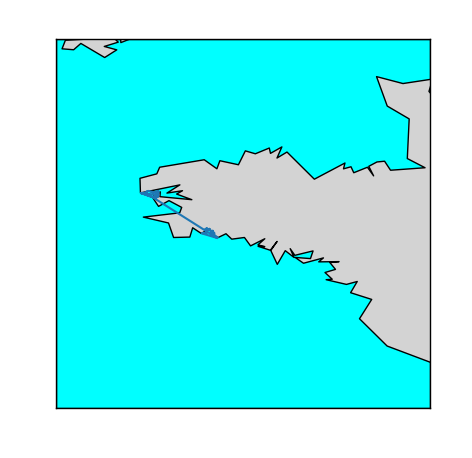


Figure 34: Fishing constraint areas

### Plot fishing areas(python)

Source code: fishing\_area/fishing.py



Figure 35:Fishing areas

### Plot ports of Brittany(python)

Source code: port\_britany/ port\_britany.py

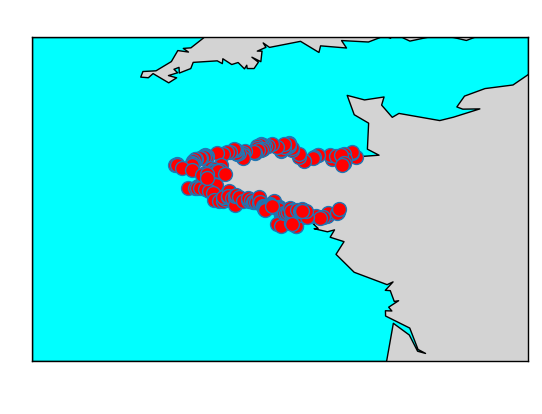


Figure 36:Ports of Brittany

### Plot the ports of the Europe (python)

Source code: world\_port\_index/wpi.py

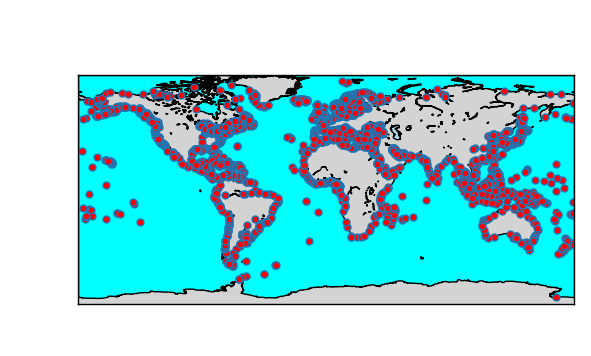


Figure 37:Ports all over the world

## 2.3. Repository

* The source code for the preliminary data analysis is located at:

<https://github.com/salevizo/cer.git>

# 3. Automata-based CER with Flink

## 3.1. Complex event detecting with automata-based framework

Since we have already understand the dataset, a system for online monitoring marine activity over streaming positions from numerous vessels sailing at sea will be presented. The system detects abnormal behavior in AIS messages emitted from vessels across time. AIS messages will be used to recognize suspicious events of one or multiple vessels.

### 3.1.1. System Architecture

The online monitoring system is a combination of the following systems: kafka and Apache Flink. Kafka is an open source streaming platform which is used for stream processing and apache Flink is a real time processing engine for stateful computations.

A Flink program is defined data streams and their transformations. A stream is a flow of events, and a transformation is an operation that takes one or more streams as input, and produces one or more output streams as an output. When executed, Flink programs are mapped to directed acyclic graphs, consisting of streams and transformation operators. Each graph starts with one or more sources and ends in one or more sinks.

### 3.1.2. System Deployment

The system implemented detects complex patterns in a stream of AIS messages. AIS messages are stored in a PSQL database. A python script has been written in order to fetch these messages from database and forward them to a kafka topic named “AIS ”.

The system consists of 2 different jobs. The 1st one is used to detect trajectory events for one vessel and the 2nd to detect complex event for more than one vessels based on the trajectory events that have been already detected at the previous step. The 1st project’s “flinkcep” input is what is written at topic AIS (the AIS messages fetched from database). The output of this project is forwarded to other kafka topics like OUT\_GAP, OUT\_COTRAVEL, OUT\_COURSE that given as input to the kafka producer of the 2nd project “cep\_flinkcep”. The results of both projects are written in txt and csv files. Each row of the txt/csv files contains information for the detected event. Csv files will be used to make some visualizations.

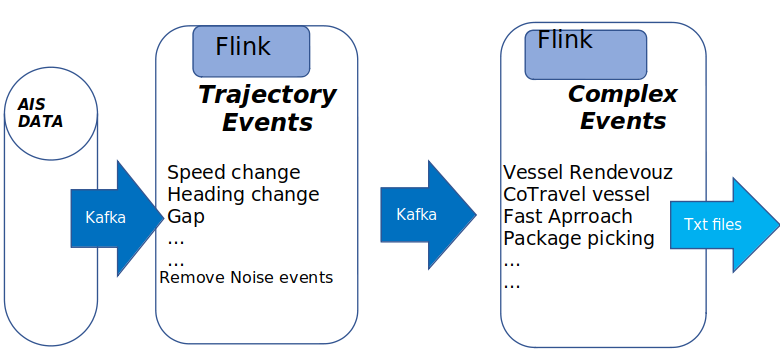


Figure 38: System deployment

### 3.1.3. Online Noise Reduction

AIS dataset, is a dataset that contains data which can be noisy, and as a result, difficult to be processed. Aiming to have better and more accurate results in the monitoring system, we should remove these noisy messages. In this implementation critical conditions have been added at the flinkcep patterns to identify noisy AIS messages, for examples emitted messages where the shift of the heading is more than 60 degrees. These conditions are checked while processing the patterns in order to improve the performance of the system.

#### Out of order events

As it is known, transmission delays may frequently occur between the original message and its arrival. Successive positional messages from a single vessel may often arrive intermingled at a distorted order. In our case, we used custom watermarks to handle the timestamp of events and we specifically decided to not accept events arriving out of order. The main reason behind the decision was that the program’s execution time was exponentially bigger if we should consider as pattern candidates the unordered events. If this system was in a real world at which we would have delays, we should take care of messages that arrived out of order by adding a small time window of accepted events despite the fact of having out of order timestamps.

### 3.1.4. Grid partitioning

We decided to translate use the coordinates of a vessel and check how far is from an area or how far is from other vessels *geohash*, a grid portioning method has been used. Geohash is a geocoding system based on a hierarchical spatial data structure which subdivides space into buckets of grid (github.com/davidmoten/geo).

Each cell is labeled using a geohash which is of user-definable precision:

● High precision geohash have a long string length and represent cells that cover only a small area.

● Low precision geohash have a short string length and represent cells that each cover a large area.

GeoHash, can have a choice of precision between 1 and 12. As a consequence of the gradual precision degradation, nearby places will often present similar prefixes. The longer a shared prefix is, the closer the two places are. In the current implementation the below cell dimensions have been used.

We preferred geohash over haversine distance because it produced must faster results for the given latitudes/longitudes.

The following table is presented as an example of the translation between geohash string length and actual distance metric units.

|  |  |
| --- | --- |
| **Precision** | **Cell dimension** |
| 4 | 39,1 km x 19,5 km |
| 5 | 4,9 km x 4,9 km |
| 6 | 1,2 km x 600,4m |
| 7 | 152,9km x 152.4m |

## 3.2. Trajectory Events

Trajectory Detection is the main module of the project since the Complex Event Recognition module uses its outcome to compute the patterns that satisfy the conditions predefined. Trajectory Movement can be vessel Speed Change Events, gap in communication and generally events that related with the movement and the behavior of one vessel or more, but mainly serve as a filter for complex events that will be checked for pattern recognition in the following project.

At this phase, system deduce instantaneousevents by examining the trace of each vessel alone. This system consumes a stream of AIS tracking messages from vessels and continuously detects important patterns that characterize their movement.

A sample AIS message that is consumed by the patterns written for trajectory detection is:

*"lat":2.541122,"lon":3.90484,"mmsi":14,"status":7,"speed":30,"turn":,"heading":36,"course":13,1, "t": 1443650402*

Trajectory detection creates a stream of important pattern of events that will be used from the complex event recognition system. Each such event is accompanied by the coordinates or some other characteristics of each corresponding vessel. At each class of trajectory events there is a message serializer class that serializes the accepted events and create another stream. These produced streams will either be used from the 2nd complex event recognition system or either they will be written on txt/csv files if there is no need to forward the stream to the 2nd monitoring system.

Below are described the trajectory events we have implemented.

### 1. Gaps in communication

System detects vessels which has communication gaps on sending AIS messages. We can define a gap as “*the absence of emitted AIS messages from a specific vessel for a specific period*”. In our approach, the ground truth of this time is 600 seconds . So, if the system won’t receive an AIS message from a vessel between 600 seconds an its far away from a port (grid 1,2 km x 600,4m), this vessel is tagged as suspicious. This pattern characterized by the critical events “gap\_start” – “gap\_end”. Information about the geohash of ports has been fetched from database.

The flinkcep pattern is written bellow. As time window we have used 900 seconds. That all the continuous AIS messages have to differ 600 seconds in order to be detected from the pattern and the message gap must not exceed the value of 900.For larger windows we faced problem with the ram (heap exceeded).

Table 1: Gap Pattern

|  |
| --- |
| **private static int gapTime=600;**  **private static int geoHashLen=6;**  **public static Pattern<AIS Message, ?> patternGap(){**  **Pattern<AIS Message, ?> rendezvouzPattern = Pattern.<AIS Message>begin("gap\_start", AfterMatchSkipStrategy.skipPastLastEvent())**  **.next("gap\_end")**  **.where(new IterativeCondition<AIS Message>()**  **@Override**  **public boolean filter(AIS Message event,Context<AIS Message> ctx) throws Exception { for (AIS Message ev : ctx.getEventsForPattern("gap\_start")) {**  **if ((event.getT() - ev.getT()) > gapTime && (event.getT() - ev.getT()) > 0**  **&& listOfPorts.contains(GeoHash.encodeHash(event.getLat(), event.getLon(), geoHashLen)) == false)**  **{return true;**  **}**  **}).within(Time.seconds(1200));** |

### 2. Vessel speed according its type

In this case, the system detects vessels which have very small or very big speed for the vessel type. Each vessel type has its own min and max speeds defined, so in any case that a vessel overcomes these limits can be considered as an event worth recognizing. We obtained and store the information about the ship type of each vessel we have created a csv file which contains all vessel types and all the mmsis for each type (these data fetched from nari\_static table on database). For the maximum and minimum speed of each vessel type, there is also a csv file that contains max and min speed per vessel type. This file created from information that is available on web.

The flinkcep pattern is written bellow. As time window we have 30 seconds . That means at the time window of 30 seconds all the AIS messages comes from one mmsi have to satisfy the following conditions in order to be detected from the pattern.

Table 2: Too slow or too fast vessel pattern

|  |
| --- |
| **public static Pattern<AIS Message, ?> patternSpeedVesselType(){**  **Pattern<AIS Message, ?> spaciousSpeedPattern = Pattern.<AIS Message>begin("speed\_spacicious\_start",**  **AfterMatchSkipStrategy.skipPastLastEvent())**  **.followedBy("speed\_spacicious\_stop")**  **.where(new IterativeCondition<AIS Message>() {**  **public boolean filter(AIS Message event, Context<AIS Message> ctx) throws Exception {**  **for (AIS Message ev : ctx.getEventsForPattern("suspicious\_heading\_start")) {**  **String mmsi\_= String.valueOf(event.getMmsi());**  **String type="";**  **for (Object o : listOfVesselsType.keySet()) {**  **if (Arrays.asList(listOfVesselsType.get(o)).contains(mmsi\_))**  **{**  **type = o.toString();**  **}**  **}**  **if (type.equals("")==false) {**  **String speed [] = listOfVesselsMaxMinSpeed.get(type);**  **if ( ((event.getSpeed()<Float.valueOf(speed[0])) ||**  **(event.getSpeed()>Float.valueOf(speed[1]))) && (event.getT() - ev.getT()) > 0) {**  **return true;**  **} else {**  **……………………**  **}).within(Time.seconds(30));** |

### 3. Adrift - Course of ground (COG) differentiates from the heading of a vessel

In this case, the system detects vessels with different heading and course. This can happen when there are extreme weather conditions and vessels are unable to follow the desired route. Heading describes the direction that a vessel is pointed at any time relative to the magnetic north pole or geographic north pole, of 511 degrees indicates noise. As such, a stationary vessel ex. a vessel which has been tied to a dock will have a heading associated with the vessel's orientation. COG describes the direction of motion with respect to the ground that a vessel has moved relative to the magnetic north pole or geographic north pole. An alert sign could be sent in this case. COG equals to 511 means unavailable gps, and these messages will be excluded. Accepted difference between heading and course over ground is every difference less than 10 degrees whereas suspicious can be considered every difference bigger than 10 and smaller than 60 degrees. At this case we also check the speed of a vessel, if the speed of the vessel is less than 1 KNOT it means that the vessel is anchored, and we do not accept those events. We care about vessels whose speed is between 1-48.6 KNOTS, this indicates that the vessel is under way. [1]

The flinkcep pattern is a simple condition and there is no window time.

Table 3: Heading and CoG pattern

|  |
| --- |
| **static int minDiff=10;**  **static int maxDiff=60;**  **static double max\_speed=48.6;**  **static double min\_speed=1;**  **public static Pattern<AIS Message, ?> patternSpaciousHeading(){**  **Pattern<AIS Message, ?> spaciousHeading = Pattern.<AIS Message>begin("suspicious\_heading\_start")**  **.where(new SimpleCondition<AIS Message>() {**  **@Override**  **public boolean filter(AIS Message event) throws Exception {**  **float courseDiffToHead=(Math.abs(event.getHeading()-event.getCourse()));**  **if (courseDiffToHead>minDiff && courseDiffToHead<maxDiff && event.getSpeed()>min\_speed**  **&& event.getSpeed()<max\_speed) { return true; …………….** |

### 4. High speed near port

At this case system detects vessels whose speed is bigger than 5KNOTS, they are near ports (grid 1,2 km x 600,4m) and at least 10 consecutive messages of a vessel satisfy those conditions. From the moment that we identify this behavior, the program does not search for intermediate patterns and follows “*skip after match* *strategy*”. Ships that overcome these speed limits can be considered as an ongoing suspicious vessel following a dangerous route. Information about the geohash of ports has been fetched from database.

The flinkcep pattern is written bellow. As time window we have 600 seconds . That means at the time window of 600 seconds all the AIS messages from one mmsi, have to satisfy the following conditions in order to be detected from the pattern.

Table 4:High Speed near port

|  |
| --- |
| **static int indexNearPorts=6;**  **static int maxSpeed=5;**  **static int patternTime=600;**  **static int patternTime=600; public static Pattern<AIS Message, ?> patternSpeedNearPort(){**  **Pattern<AIS Message, ?> fastForwardPattern = Pattern.<AIS Message>begin("acceleration\_start", AfterMatchSkipStrategy.skipPastLastEvent())**  **.where(new SimpleCondition<AIS Message>() {**  **@Override**  **public boolean filter(AIS Message event) throws Exception {**  **if (listOfPorts.contains(GeoHash.encodeHash(event.getLat(), event.getLon(), indexNearPorts))== true**  **&& event.getSpeed()>maxSpeed) {**  **return true;**  **}return false;**  **}}).times(10)**  **.consecutive()**  **.within(Time.seconds(patternTime));** |

### 5. Long Term Stop

Long-Term stop is only triggered if the vessel is noticed to move slowly after a pause. Pause means that vessel’s speed is less than 1KNOT. [1] If the next M messages, within 200 seconds are inside a predefined area (grid 1,2 km x 600,4m) , a long term stop is identified. More specifically, this pattern first detects vessels whose speed is smaller than 1KNOT and far away from ports (grid 1,2 km x 600,4m). After check if the detected vessels don’t change their speed more than 2 KNOTS within 1800 seconds and they send AIS messages inside an area (grid 1,2 km x 600,4m), means they haven’t moved away. Information about the geohash of ports has been fetched from database.

The flinkcep pattern is written bellow. As time window we have 7200 seconds . That means at the time window of 7200 seconds all the AIS messages from one mmsi have to satisfy the following conditions in order to be detected from the pattern.

Table 5: Long Term Stop

|  |
| --- |
| **public static Pattern<AIS Message, ?> patternLongStop() {**  **int longterm = 3600;**  **int time\_window=7200;**  **Pattern<AIS Message, ?> LongTermStopPattern = Pattern.<AIS Message>begin("start", AfterMatchSkipStrategy.skipPastLastEvent())**  **.next("stop")**  **.where(new SimpleCondition<AIS Message>() {**  **@Override**  **public boolean filter(AIS Message event) throws Exception {**  **if((event.getSpeed() <1 && listOfPorts.contains(GeoHash.encodeHash(event.getLat(), event.getLon(), 8)) == false)){**  **String geoHash1=GeoHash.encodeHash(event.getLat(),event.getLon(),8);**  **return true;**  **}**  **return false;**  **}**  **})**  **.followedBy("stop\_ends")**  **.where(new IterativeCondition<AIS Message>() {**  **@Override**  **public boolean filter(AIS Message event, Context<AIS Message> ctx) throws Exception {**  **for (AIS Message ev : ctx.getEventsForPattern("stop")) {**  **if ( ev.getMmsi() == event.getMmsi()) {**  **String geoHash1=GeoHash.encodeHash(event.getLat(),event.getLon(),6);**  **String geoHash2=GeoHash.encodeHash(ev.getLat(),ev.getLon(),6);**  **if((geoHash1.equals(geoHash2)) && (event.getSpeed()<2) && ev.getSpeed()<2 && (Math.abs(event.getT() - ev.getT()) > 1800)){**  **return true;**  **}**  **else{**  **return false;**  **}**  **}.within(Time.seconds(time\_window));**  **…..**  **return false; }})** |

### 6. Vessels with false status

The dataset contains vessels with wrong status for example vessels with speed more than 5KNOTS and status moored or at anchor (status =1, status=5). Those ships could be identified as noise.

The flinkcep pattern is a simple condition and there is no window time.

Table 6: Vessels with false status

|  |
| --- |
| **public static Pattern<AIS Message, ?> patternFalseType() {**  **Pattern<AIS Message, AIS Message> alarmPattern = Pattern.<AIS Message>begin("first")**  **.where(new SimpleCondition<AIS Message>() {**  **@Override**  **public boolean filter(AIS Message event) {**  **if ((event.getStatus() == 1 || event.getStatus() == 5)) {**  **if (event.getSpeed() > 5) {**  **return true;**  **} else**  **return false;**  **…………………….. });** |

## 3.3. Complex event

Complex event Recognition module consumes the output of the Trajectory detection module process the results and recognize in real time potentially complex maritime situations. Below will be described some patterns used for complex event recognition. For some of them, two monitoring systems needed. The 1st system for the trajectory events that are analyzed at the previous section and the 2nd one for combining these events to detect more complex patterns.

### 1. Two vessels co-travelling

The pattern, checks that the AIS messages from 2 different vessels are between a time period of 15 seconds . The pattern will detect events that happened closely in the time dimension. Next thing that will be checked is if the ships are on route and not paused. The speed of the vessel should be bigger than 1KNOT (trajectory event - 1 KNOT is the minimum speed of a vessel that isn’t in anchor). [1] At the end, the geohash of the 2 vessels will be checked, ensuring by that, that the two vessels were located in the same grid. The precision of geohash is 6 (grid: 1.2 km x 600m). Two vessels seem to cotravel if they are close enough (grid: 1.2 km x 600m) for more than 180 seconds.This pattern characterized from events “vessel\_1” – “vessel\_2”.

In the end, all the detected patterns will be inserted into a kafka topic and will be given as input stream at the 2nd online monitoring system “cep\_flincep” to detect the desired events.

In order to make the second job more efficient, we decided to insert the id of the vessel that is the smallest between mmsi 1 and 2 as MMSI\_1 and to collect all the events per vessel. With that, instead of search through all the accepted events we will consider only the pairs that at least one of the MMSI does not change.

As time window we have used 35 seconds for the 1st pattern and 300 seconds for the 2nd. 35 seconds of the 1st pattern means all the AIS messages (for all vessels) in a time window of 35 seconds should be at the same geohash in order to be detected from the system. The second pattern checks that vessels emitted from the first pattern, follow the following rules:

* Two different messages for two different ships that co-exist in the same grid must at worst be differentiated by a margin of 35 seconds.
* Going backwards to the accepted messages following the first rule, we check whether those messages surpass in total the 180 seconds margin. Base event is considered the latest, and we move iteratively backwards until we find a series of messages that the difference between an event and the base event is more than 180 seconds.
* All this should happen between 300 seconds.

The 1st pattern is :

Table 7: Cotravel for 2 vessels pattern – trajectory events

|  |
| --- |
| **static int minSpeed=1;**  **static int geoHashLength=6;**  **static int timeBetweenVesselsMsgs=15;**  **public static Pattern<AIS Message, ?> patternCoTravel(){**  **Pattern<AIS Message, ?> coTravelPattern = Pattern.<AIS Message>begin("vessel\_1")**  **.subtype(AIS Message.class)**  **.followedBy("vessel\_2")**  **.subtype(AIS Message.class)**  **.where(new IterativeCondition<AIS Message>() {**  **@Override**  **public boolean filter(AIS Message event, Context<AIS Message> ctx) throws Exception {**  **for (AIS Message ev : ctx.getEventsForPattern("vessel\_1")) {**  **if(ev.getSpeed()>minSpeed**  **&& event.getSpeed()>minSpeed**  **&& ev.getMmsi()!=event.getMmsi()){**  **String geoHash1=**  **GeoHash.encodeHash(ev.getLat(),ev.getLon(),geoHashLength);**  **String geoHash2=**  **GeoHash.encodeHash(event.getLat(),event.getLon(),geoHashLength);**  **if(geoHash1.equals(geoHash2)==true){**  **return true;**  **}else{return false;**  **}……………**  **.within(Time.seconds(timeBetweenVesselsMsgs));** |

The 2nd pattern is:

Table 8: Cotravel for 2 vessels pattern – complex events

|  |
| --- |
| **static int coTravelTime=35;**  **static int coTravellingTotalTime=180;**  **Pattern<CoTravelInfo, ?> coTravelattern = Pattern.<CoTravelInfo>begin("msg\_1",AfterMatchSkipStrategy.skipPastLastEvent())**  **.subtype(CoTravelInfo.class)**  **.oneOrMore()**  **.followedBy("msg\_2")**  **.where(new IterativeCondition<CoTravelInfo>() {**  **@Override**  **public boolean filter(CoTravelInfo event, Context<CoTravelInfo> ctx) throws Exception{**  **int base = event.getTimestamp();**  **int currTime = event.getTimestamp();**  **List<CoTravelInfo> l = Lists.newArrayList(ctx.getEventsForPattern("msg\_1"));**  **for (CoTravelInfo ev : Lists.reverse(l)) {**  **if ((currTime - ev.getTimestamp()) < coTravelTime) {**  **if (event.getMmsi\_2() == ev.getMmsi\_2()) {**  **if ((base - ev.getTimestamp()) > coTravellingTotalTime) {**  **return true;**  **} else {**  **currTime = ev.getTimestamp();**  **}}} else {**  **return false;**  **}………….**  **.within(Time.seconds(300));** |

### 2. Fishing Activity

This pattern combines two trajectory events in order to detect a complex event. Specifically, it checks the sequence of continuous changes of the heading (changes between 15 and 60 degrees) , followed by a gap in communication (600 seconds ) and a final turn in the end (change between 15 and 60 degrees). This can be considered as an alert sign for illegal fishing. The captain checks the area and afterwards closes its GPS. At that point the system will check just the events that have been characterized by the tag “gap\_start” – “gap\_end” in order to detect some more changes in heading of the vessels. The trajectory events of heading change and gap in communication have been already analyzed.

The flinkcep pattern is written bellow. As time window we have used 1200 seconds . That means at the time window of 1200 seconds the AIS messages from one mmsi have to satisfy the following conditions in order to be detected from the pattern.

Table 9: Fishing pattern

|  |
| --- |
| **static int headingChangeMin=15;**  **static int headingChangeMax=60;**  **static int gapTime=600;**  **public static Pattern<AIS Message, ?> patternFishing(){**  **Pattern<AIS Message, ?> fishingPattern = Pattern.<AIS Message>begin("start")**  **.subtype(AIS Message.class)**  **.followedBy("gap\_start")**  **.subtype(AIS Message.class)**  **.where(new IterativeCondition<AIS Message>() {**  **@Override**  **public boolean filter(AIS Message event, Context<AIS Message> ctx) throws Exception {**  **for (AIS Message ev : ctx.getEventsForPattern("start")) {**  **if(Math.abs(ev.getHeading()-vent.getHeading())>headingChange)**  **{return true;**  **}else{**  **return false;…… }})**  **.subtype(AIS Message.class)**  **.followedBy("gap\_end")**  **.subtype(AIS Message.class)**  **.where(new IterativeCondition<AIS Message>() {**  **@Override**  **public boolean filter(AIS Message event, Context<AIS Message> ctx) throws Exception {**  **for (AIS Message ev : ctx.getEventsForPattern("gap\_start")) {**  **if((event.getT()-ev.getT())>gapTime){**  **return true;**  **}else{**  **return false; }}**  **return false;}})**  **.followedBy("change in heading again")**  **.subtype(AIS Message.class)**  **.where(new IterativeCondition<AIS Message>() {**  **@Override**  **public boolean filter(AIS Message event, Context<AIS Message> ctx) throws Exception {**  **for (AIS Message ev : ctx.getEventsForPattern("gap\_end")) { if(Math.abs(ev.getHeading()-event.getHeading())>headingChange){**  **return true;**  **}else{**  **return false; }}**  **return false;}})**  **.within(Time.seconds(600));** |

### 3. Vessel Rendezvous

The pattern, checks that the AIS messages from 2 different vessels if they have gap in their communication are in the same geohash grid. The precision of geohash is 6 (grid: 1,2 km x 600,4m). The information about the geohash of each vessel is given by the previous pattern “gap-communication” whose outcome is used as input stream ain this pattern. Two vessels may have arranged a rendezvous while on the sea, when they have both gap in their communication at the same time as long as they are at the same geohash grid.

This pattern characterized from events “vessel\_1” – “vessel\_2”.

The flinkcep patterns are written bellow. As time window we have used 900 seconds for the 1st pattern and 120 seconds for the 2nd. 900 seconds of the 1st pattern means all the continuous AIS messages of a mmsi have to differ 600 seconds in order to be detected. With the second pattern, we aim to identify vessels that their gap end event differs at most to 120 seconds.

The 1st pattern is:

Table 10: Vessel Rendezvous pattern - trajectory events

|  |
| --- |
| **private static int gapTime=600;**  **private static int geoHashLen=6;**  **public static Pattern<AIS Message, ?> patternGap(){**  **Pattern<AIS Message, ?> rendezvouzPattern = Pattern.<AIS Message>begin("gap\_start", AfterMatchSkipStrategy.skipPastLastEvent())**  **.followedBy("gap\_end")**  **.where(new IterativeCondition<AIS Message>()**  **@Override**  **public boolean filter(AIS Message event,Context<AIS Message> ctx) throws Exception { for (AIS Message ev : ctx.getEventsForPattern("gap\_start")) {**  **if ((event.getT() - ev.getT()) > gapTime && (event.getT() - ev.getT()) > 0**  **&& listOfPorts.contains(GeoHash.encodeHash(event.getLat(), event.getLon(), geoHashLen)) == false)**  **{**  **return true;**  **}**  **………….**  **}).within(Time.seconds(1200));** |

The 2nd pattern is:

Table 11: Vessel Rendezvous pattern - complex events

|  |
| --- |
| **public static int gapTime=120;**  **public static Pattern<GapMessage, ?> patternRendezvouz(){**  **Pattern<GapMessage, ?> rendevouzPattern = Pattern.<GapMessage>begin("Vessel\_1")**  **.subtype(GapMessage.class)**  **.followedBy("Vessel\_2")**  **.subtype(GapMessage.class)**  **.where(new IterativeCondition<GapMessage>() {**  **@Override**  **public boolean filter(GapMessage event, Context<GapMessage> ctx) throws Exception {**  **for (GapMessage ev : ctx.getEventsForPattern("Vessel\_1")) {**  **if ((ev.getGeoHash().equals(event.getGeoHash()) == true)**  **&& ev.getMmsi()!=event.getMmsi()) {**  **return true;**  **} else { return false;**  **}}**  **return false;}})**  **.within(Time.seconds(gapTime));** |

### 4. Adrift

The pattern, detects vessels that may have a problematic route. The vessels checked at this pattern are already detected as vessels with suspicious difference between heading and course of ground.

As far as the second pattern is concerned, the following prerequisites must be met:

* Two consecutive messages for a vessel must be on a time range of 20 seconds.
* Going backwards to the accepted messages following the first rule, we check whether those messages surpass in total the 180 seconds margin. Base event is considered the latest, and we move iteratively backwards until we find a series of messages that the difference between an event and the base event is more than 120 seconds.
* All this should happen between 240 seconds.

The 1st pattern is:

Table 12: Big difference between heading and course – trajectory event

|  |
| --- |
| **static int minDiff=10;**  **static int maxDiff=60;**  **static double max\_speed=48.6;**  **static double min\_speed=1;**  **public static Pattern<AIS Message, ?> patternSpaciousHeading(){**  **Pattern<AIS Message, ?> spaciousHeading = Pattern.<AIS Message>begin("suspicious\_heading\_start")**  **.where(new SimpleCondition<AIS Message>() {**  **@Override**  **public boolean filter(AIS Message event) throws Exception {**  **float courseDiffToHead=(Math.abs(event.getHeading()-event.getCourse()));**  **if (courseDiffToHead>minDiff && courseDiffToHead<maxDiff && event.getSpeed()>min\_speed**  **&& event.getSpeed()<max\_speed) {**  **return true;**  **} else {**  **return false; …………….** |

The 2nd pattern is:

Table 13: Problematic route - complex events

|  |
| --- |
| **static int messagesTimeGap=20;**  **static int problematicRouteTime=120;**  **public static Pattern<courseHead, ?> patternSpaciousHeading(){**  **Pattern<courseHead, ?> spaciousHeading = Pattern.<courseHead>begin("suspicious\_heading\_start", AfterMatchSkipStrategy.skipPastLastEvent())**  **.oneOrMore()**  **.followedBy("suspicious\_heading\_stop")**  **.where(new IterativeCondition<courseHead>() {**  **@Override**  **public boolean filter(courseHead event, Context<courseHead> ctx) throws Exception {**  **int base = event.getTimestamp();**  **int currTime = event.getTimestamp();**  **List<courseHead> l = Lists.newArrayList(ctx.getEventsForPattern("suspicious\_heading\_start"));**  **for (courseHead ev : Lists.reverse(l)) {**  **if ((currTime - ev.getTimestamp()) < messagesTimeGap) {**  **if ((base - ev.getTimestamp()) > problematicRouteTime) {**  **return true;**  **} else {**  **currTime = ev.getTimestamp();**  **}} else {**  **return false; ……..**  **}).within(Time.seconds(240));** |

### 5. Package Picking

This pattern detects possible interaction between two vessels. A possible interaction is when one of them drops a package at some area and another vessel appears later in order to pick it up. By joining the previous long stop events using geohash area (grid: 1,2 km x 600,4m) we find ships that have long stops in the same area where the package picking is possible. As described before the ships should be away from ports. The timestamp between the vessels to be in the same geohash area is 60seconds .

The flinkcep patterns are written bellow. As time window we have used 3600 seconds . That means at the time window of 3600 seconds the AIS messages satisfy the following conditions in order to be detected from the pattern.

|  |
| --- |
| **Pattern<SuspiciousLongStop, ?> alarmPattern = Pattern.<SuspiciousLongStop>begin("first")**  **.where(new SimpleCondition<SuspiciousLongStop>() {**  **@Override**  **public boolean filter(SuspiciousLongStop suspiciousLongStop) throws Exception {**  **if(suspiciousLongStop.getMmsi()>0) {**  **return true; }else{**  **return false; }}})**  **.followedBy("picking")**  **.where(new IterativeCondition<SuspiciousLongStop>() {**  **@Override**  **public boolean filter(SuspiciousLongStop event, Context<SuspiciousLongStop> ctx) throws Exception {**  **String geoHash1= GeoHash.encodeHash(event.getLat(),event.getLon(),6);**  **for (SuspiciousLongStop ev : ctx.getEventsForPattern("first")) {**  **if ( ev.getMmsi() != event.getMmsi()) {**  **String geoHash2=GeoHash.encodeHash(ev.getLat(),ev.getLon(),6);**  **if((geoHash1.equals(geoHash2)) && (event.getGapEnd() - ev.getGapEnd())<60){**  **return true; }**  **else{return false; }}) ……..**  **.within(Time.seconds(3600));** |

### 6. Loitering

Loitering is the act of remaining in a particular area for a long period without purpose. Vessels with low speed, anchored or moored must be filtered out. If the messages from a single vessel are in the same area for 1800 seconds (loitering time , Ltrtime) this vessel is considered to be loitering. So firstly detects vessels that are far away from ports (grid: 1,2 km x 600,4m) with speed bigger than 2.87 KNOTS and smaller than 8 KNOTS and after check id this vessel remain at the same area. Information about the geohash of ports has been fetched from database. [1]

The flinkcep patterns are written bellow. No pattern window is applied we simply search for vessels for loitering. Optional time window would be 3600 (1 hour)

|  |
| --- |
| **Pattern<AIS Message, ?> Loitering = Pattern.<AIS Message>begin("stop")**  **.subtype(AIS Message.class)**  **.where(new SimpleCondition<AIS Message>() {**  **@Override**  **public boolean filter(AIS Message event) throws Exception {**  **boolean near\_ports = false;**  **for(String str: Ports) {**  **String ship\_geohash = GeoHash.encodeHash(event.getLat(),event.getLon(),6);**  **if(str.equals(ship\_geohash))**  **near\_ports = true;**  **}**  **if((event.getSpeed() >2.87 && near\_ports == false)){**  **if(event.getSpeed() < 8){**  **String geohash1=GeoHash.encodeHash(event.getLat(),event.getLon(),6);**  **return true;}**  **else{**  **return false;**  **…………………………. })**  **.followedByAny("stop\_ends")**  **.where(new IterativeCondition<AIS Message>() {**  **@Override**  **public boolean filter(AIS Message event, Context<AIS Message> ctx) throws Exception {**  **for (AIS Message ev : ctx.getEventsForPattern("stop")) {**  **String geoHash1=GeoHash.*encodeHash*(event.getLat(),event.getLon(),6)**  **if ( ev.getMmsi() == event.getMmsi()) {**  **String geoHash2=GeoHash.*encodeHash*(ev.getLat(),ev.getLon(),6);**  **if((geoHash1.equals(geoHash2)) && ev.getSpeed()< 8 && (event.getSpeed()< 8 && (event.getT()-ev.getT()>ltrtime))){**  **if(event.getSpeed()>2.87 && ev.getSpeed() > 2.87){**  **return true;**  **}** |

## 3.4. Producer of AIS messages for the kafka topic

The AIS messages are fetched from database with a python script and after given as input at a kafka topic. The python script is inside the path flinkcep/producer and the script is: AIS .py. Running the script kafka topic’s name should be defined.

## 3.5. Watermark Pattern

We take care of the time in our events by introducing the watermark functionality. Watermark is the assigning of a timestamp into an event by using one of its fields instead of using the time that the event was consumed by the system.

Thus, the following implementation is made to retrieve and assign the timestamp of the ais message to the event.

We also extended this functionality to the second project. We wanted also there to have events based on event time and not on the system time. The only variation with respect to the first project is that we use a different field for time extraction.

Table 14: Watermark pattern

|  |
| --- |
| ***DataStream<AIS Message> messageStream = env .addSource(new FlinkKafkaConsumer09<>(***  ***parameterTool.getRequired("INPUT"),new AIS MessageDeserializer(),props))***  ***.assignTimestampsAndWatermarks(new Watermarks());*** |

## 3.6. Running Apache Flink Jobs

As analyzed before system consists of 2 projects. For each project there is a different job running on flinkcep. For each job there are different patterns that are checked. The following screenshots show a sample execution of them. The system is running for 1h 48 min, that means system receives AIS messages for 1h 48 min. During this period of time, system detects trajectory and complex events.

You can check how long the job runs as long as the number of events which has been detected.

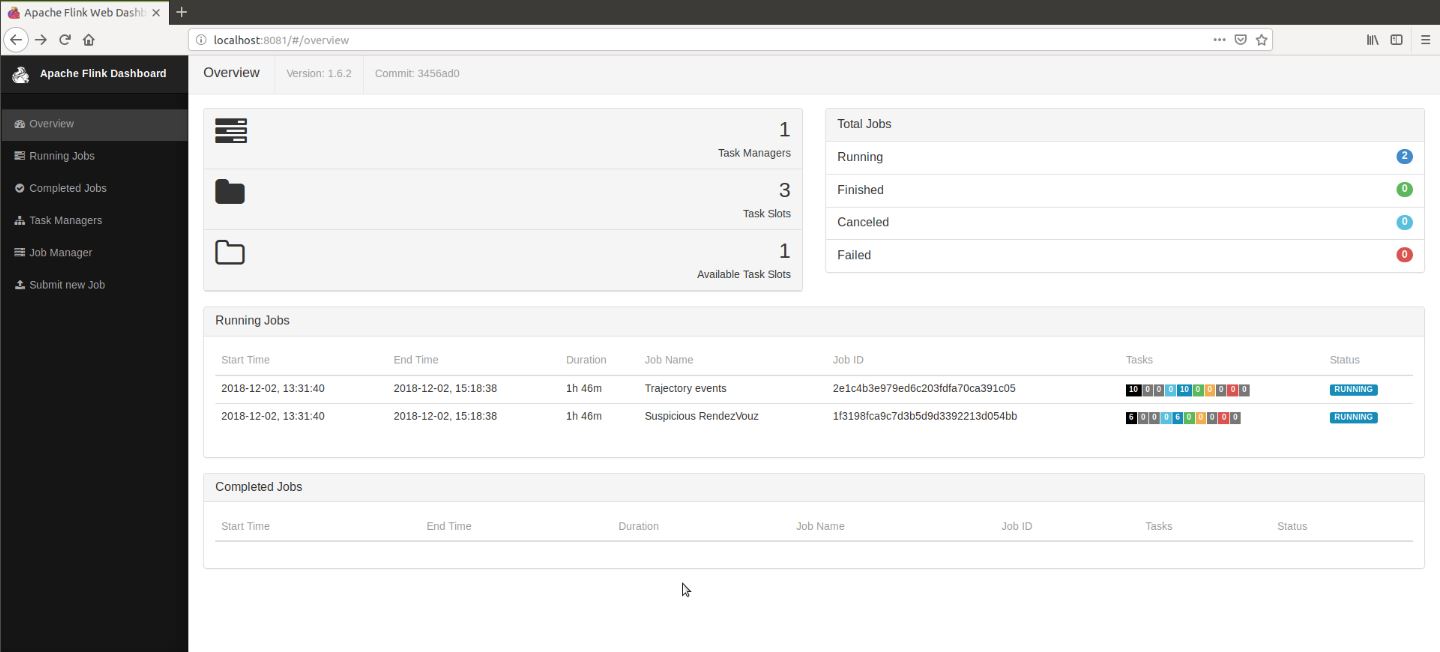


Figure 39:Two jobs are running: trajectory and complex events



Figure 40: Patterns for job Id: trajectory event (part a)

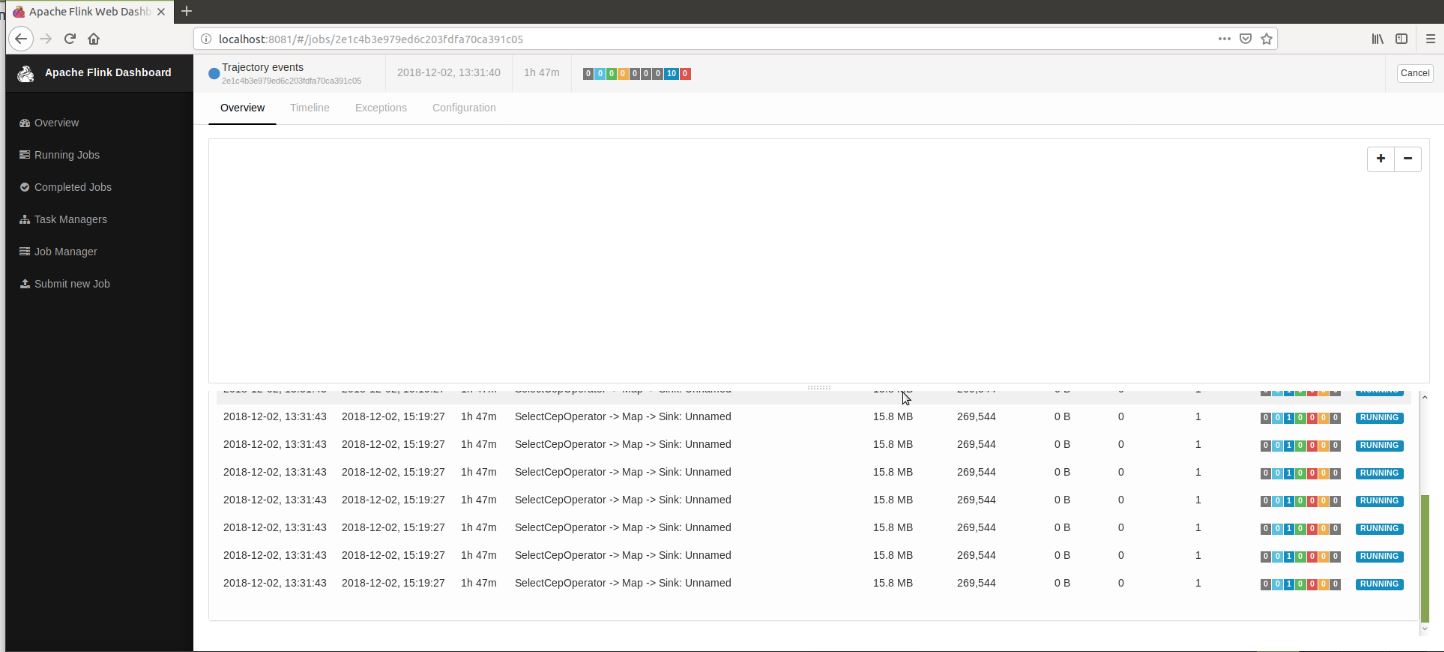


Figure 41:Patterns for job Id: trajectory event (part b)



Figure 42: Patterns for job Id: complex event

At this job, which takes as input the output of the 1st job the acyclic graphs are shown, consisting of streams and transformation operators. Each graph starts with one source and end in one sink.

## 3.7. Outcomes of the running jobs

All the above patterns, while running they write their results (information about the detected events) into .txt files at the path /Desktop/temp. After that, some python scripts will be executed (at the path /conver\_txt\_to\_csv ) for parsing these .txt files and make .csv files which contain just the values of fields that are needed to make visualizations on qgis.

The .txt files are listed below:

* speed\_near\_port.txt, fishing.txt, false\_speed.txt, loitering.txt, longstop.txt, packages.txt, rendezvous.txt, cotravel.txt, coursehead.txt, suspicious\_speed.txt

The scripts used for these files are:

|  |  |
| --- | --- |
| files | script |
| coursehead.txt | decode.py |
| cotravel.txt | decode5.py |
| loitering.txt | decode1.py |
| Packages.txt | decode2.py |
| rendezvous.txt | decode3.py |
| speed\_near\_port.txt | decode6.py |
| fishing.txt | decode4.py |

## 3.8. Empirical Evaluation

Changing the values on mobility tracking parameters eg: seconds of gap in communication, seconds of gap in a suspected fishing etc we can see that the number of events that are detected changes dramatically. Doing some tests with 1.000.000 AIS messages, we conclude at the below results. Due to the fact that some patterns required excessive amount of memory, affected the pattern processing and slow down the whole process. The selection below are for patterns whose output was emitted in a reasonable time for the above amount of AIS messages. We decided to use as parameters values these that were giving us more realistic results (the bold values).

The below results came after running the monitoring system for **1.000.000** ais messages and the time period of ['1443650402','1444660250'].

### Rendezvous

|  |  |
| --- | --- |
| GeoHash | Detected events |
| 5  *4.9km x 4.9km* | 476 |
| 6  ***1.2km x 609.4m*** | 432 |
| **7**  *152m x 152m* | **195** |
| 8  *38.2m x 19m* | 58 |

### Illegal Fishing

|  |  |
| --- | --- |
| Seconds of gap in a suspected fishing activity(max 1200) | Detected illegal fishing events |
| 300 | 16 |
| **600** | **10** |
| 900 | 9 |
| 1200 | 8 |

### Speed near port

|  |  |
| --- | --- |
| GeoHash for near port | Detected events |
| 5  *4.9km x 4.9km* | 1545 |
| **6**  ***1.2km x 609.4m*** | **41** |
| 7  *152m x 152m* | 0 |
| 8  *38.2m x 19m* | 0 |

## 3.9. Visualization of the detected events

The below pictures come from QGIS framework using the .csv files with the several complex or trajectory events. These visualizations represent the detected patterns using the bold empirical values on patterns (see previous Section). These are also the values that are written at the patterns’ description on this report.

### 1. Cotraveling activity

The involved mmsis at this plot are 7: 227730220,227005550, 28051000,356101000,227004390,235095836,244740921 and the pattern is the one described at the complex event section.



Figure 43: Cotraveling activity (a)



Figure 44: Cotraveling activity with zoom

### 2. Loitering activity

1. The involved mmsi at this plot is 1: 228037700 and the pattern is the one described at the section of trajectory events.

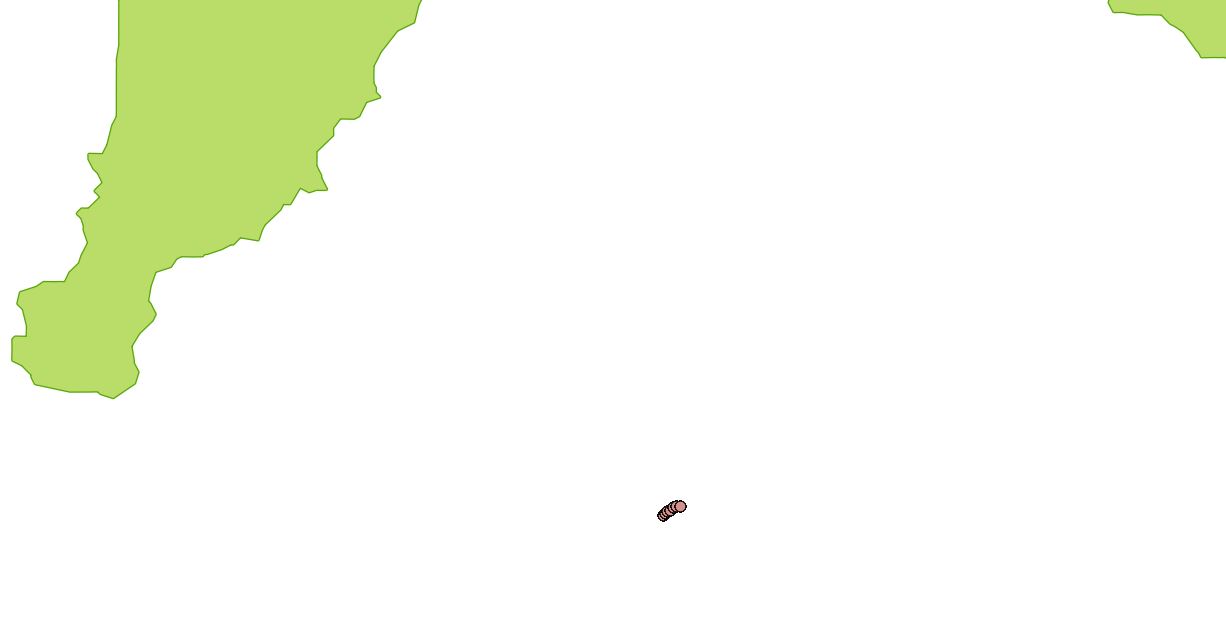


Figure 45: Loitering activity

### 3. Adrift activity

1. The involved mmsi at this plot is 1: 228064900 and the pattern is the one described at the complex events section.



Figure 46: Adrift activity

### 4. High speed near port

1. The first plot has as geohash index 6 (1.2 km x 609.4 m near port) and the involved mmsis are 6: 227577000, 227002330,228051000, 246254000,227730220,227005550. The detected events at this case are 41.

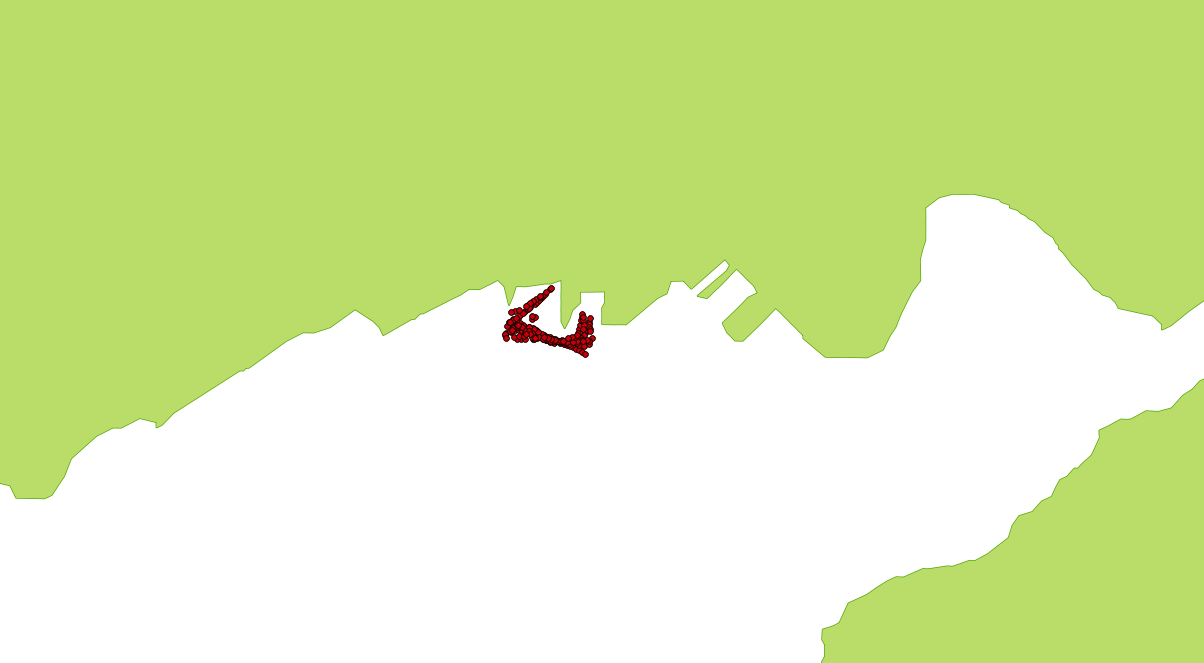


Figure 47: High speed near port, grid of 1.2km x 609.4m

1. The second plot has as geohash index 5 (4.9 km x 4.9km near port) and the involved mmsis are 108. The detected events at this case are 1545.



Figure 48: High speed near port, grid of 4.9km x 4.9km (1545 detected events)

As we can see the vessels that detected at 2nd case are also detected at the 1st. Using bigger grid as accepted distance more vessels will be detected. Events that have been detected in distance\_a are also detected in distance\_b where distance\_a<= distance \_b. We decided to choose the 1st case.

### 5. Fishing activity

The blue points on the plot used to describe the case of 300 secs as gap of communication and the involved mmsis are 11: ['228929000', '227003050', '228160000', '226177000', '228017700', '228213700', '227142200', '227318010', '227008170', '228064900', '227300000']. The detected events at this case are 16.

1. The purple points on the plot used to describe the case of 600 seconds as gap in communication and the involved mmsis are 7: ['227003050', '228160000', '228017700', '228213700', '227142200', '227318010', '227008170']. The detected events at this case are 10.
2. We can see that the 10 detected events on the 2nd case (600 secs) are also detected in the 1st case (300 secs – 16 events).

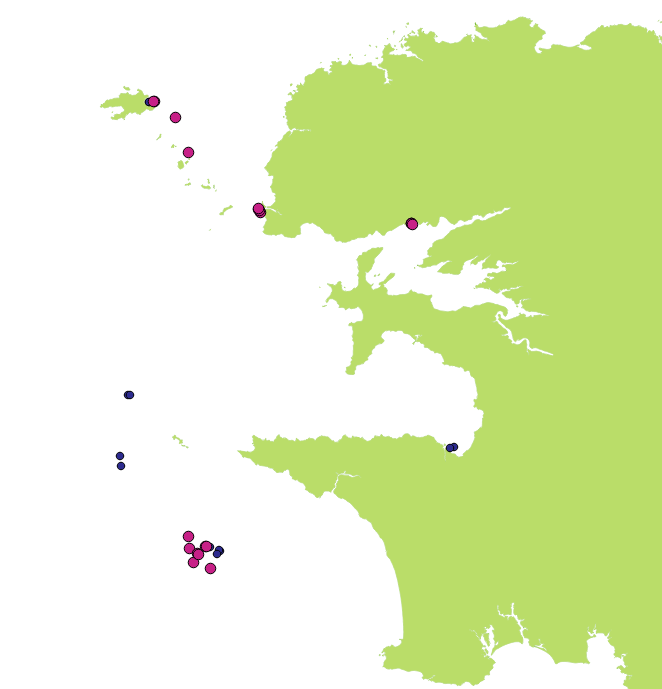


Figure 49: Fishing activity, 300 – 600 seconds gap in communication (blue:300 secs, purle:600 secs)

The black points on the plot used to describe the case of 900 secs as gap of communication and the involved mmsis are 6: ['227003050', '228160000', '228017700', '227142200', '227318010', '227008170']. The detected events at this case are 9.

The purple points on the plot used to describe the case of 600 secs gap in communication (10 detected events).

We can see that the 9 detected events on the 2nd case (900 secs) are also detected in the 1st case (600 secs – 10 events).

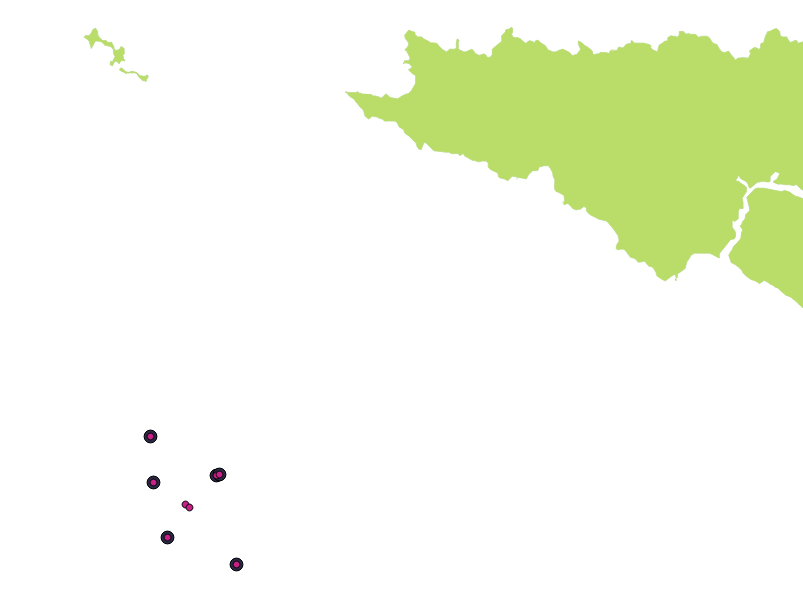


Figure 50: Fishing activity, 600 – 900 seconds gap in communication (purple:600 secs, black:900 secs)

1. The green points on the plot used to describe the case of 1200 secs of gap in communication and the involved mmsis are 5:[ '228160000', '227008170', '227003050', '227318010', '228017700']. The detected events at this case are 8.

The black points on the plot used to describe the case of 900 secs of gap in communication (9 detected events).

We can see that the 8 detected events on the 2nd case (1200 secs) are also detected in the 1st case (900 secs – 9 events).



Figure 51: Fishing activity, 900 - 1200 seconds gap in communication (green:1200 secs, black:900 secs)

1. In other words, increasing the seconds, the pattern detects less events. Events that have been detected in sec\_b are also detected in sec\_a where sec\_a<=sec\_b. We decided to use 600 secs (2nd case).

### 6. Vessels rendezvous

The blue points on the plot used to describe the case of 4.9km x 4.9km as grid and the involved mmsis are 209. The detected events at this case are 476.

1. The purple points on the plot used to describe the case of 1.2km x 609.4m as grid and the involved mmsis are 179. The detected events at this case are 432.

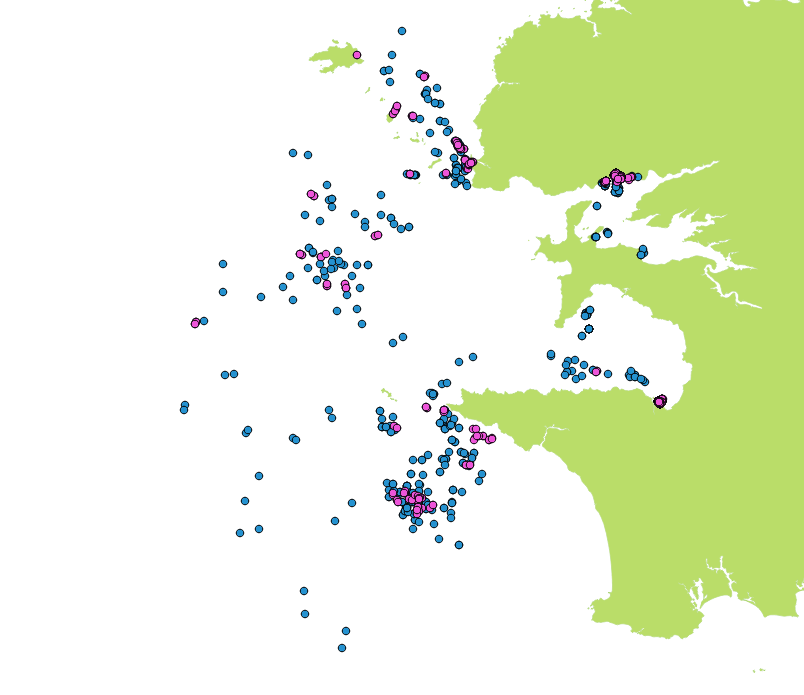


Figure 52: Vessels Rendezvous (blue: grid of 4.9km x 4.9km , purple: grid 1.2km x 609.4m)

The black points on the plot used to describe the case of 152m x 152m as grid and the involved mmsis are 101. The detected events at this case are 195.

The purple points on the plot used to describe the case of 1.2km x 609.4m as grid (432 detected events).

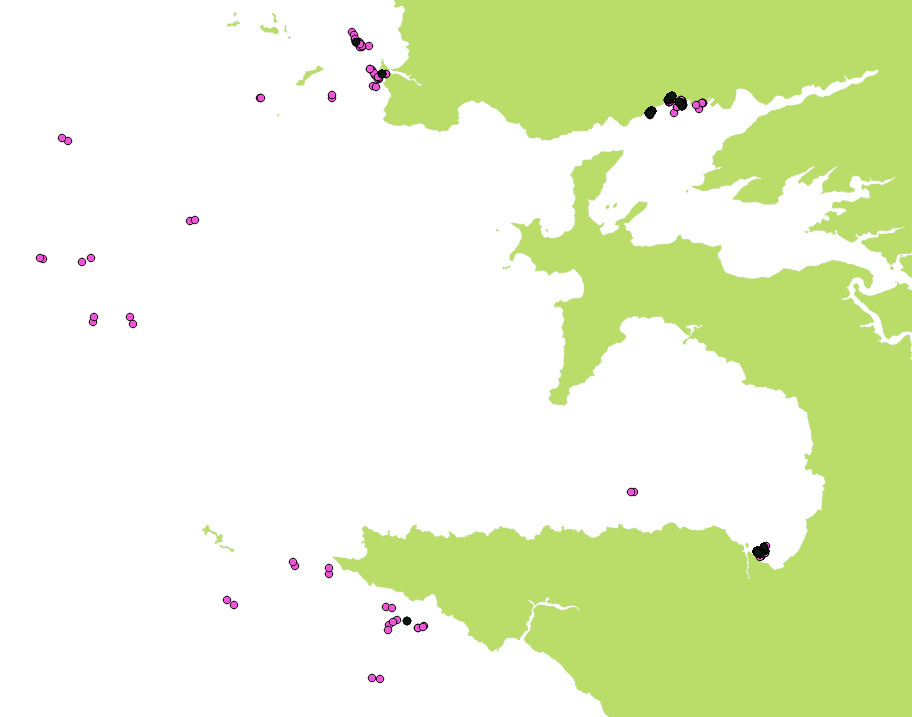


Figure 53: Vessels Rendezvous (black: grid of 152 m x 152m , purple: grid 1.2km x 609.4m)



Figure 54: Vessels Rendezvous (black: grid of 152 m x 152m , purple: grid 1.2km x 609.4m) with zoom

The yellow points on the plot used to describe the case of 38.2m x 19m as grid and the involved mmsis are 40. The detected events at this case are 58.

The black points on the plot used to describe the case of 152m x 152m as grid (195 detected events).



Figure 55:Vessels Rendezvous (black: grid of 152 m x 152m , yellow: grid 38,2 m x 19m)



Figure 56:Vessels Rendezvous (black: grid of 152 m x 152m , yellow: grid 38,2 m x 19m) with zoom

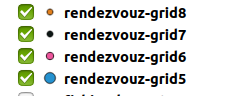
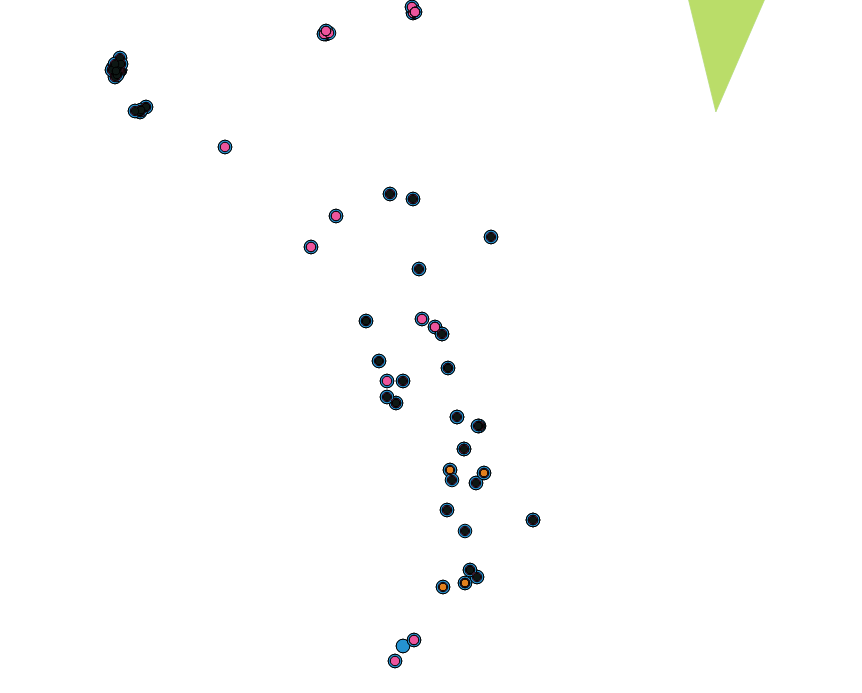
 

Figure 57:Vessels Rendezvous with zoom for all different grid values

## 3.10. Time of execution

The monitoring of about 1000000 ais messages and the detection of complex events based on the patterns described on the Sections took more than 1h.

## 3.11. Running commands

First you need to run 3 servers: flink cluster server, zookeeper and kafka server. The next step is to create 4 topics on kafka. Also, there are 2 java projects that will run and one python script which will be used as producer for the main topic of the system (the one contains AIS messages).

The commands you need to execute are described below:

* ***$ flink\_1.6.2:bin/start-cluster.sh***
* start the cluster of flink
* Check that server is running on http://localhost:8081/#/overview
* ***$ kafka2.2:bin/zookeeper-server-start.sh config/zookeeper.properties***
* ***$ kafka2.2:bin/kafka-server-start.sh config/server.properties***
* Start the kafka and zookeeper servers
* ***$ kafka2.2:bin/kafka-topics.sh --create --zookeeper localhost:2181 --replication-factor 1 partitions 1 --topic AIS***
* ***$kafka2.2:bin/kafka-topics.sh --create --zookeeper localhost:2181 --replication-factor 1 --partitions 1 --topic OUT\_GAP***
* ***$kafka2.2:bin/kafka-topics.sh --create --zookeeper localhost:2181 --replication-factor 1 --partitions 1 --topic OUT\_COTRAVEL***
* ***$kafka2.2:bin/kafka-topics.sh --create --zookeeper localhost:2181 --replication-factor 1 --partitions 1 --topic OUT\_COURSE***
  + Create the topics on kafka
* ***sudo /home/cer/Desktop/flink-1.6.2/bin/flink run /home/cer/Desktop/cer/flinkcep/cep\_flinkcep/target/flinkicu\_cep-1.0-jar-with-dependencies.jar --IN\_GAP GAP --bootstrap.servers localhost:9092 --zookeeper.connect localhost:2181 --IN\_COTRAVEL COTRAVEL --IN\_COURSE COURSE***
* ***sudo /home/cer/Desktop/flink-1.6.2/bin/flink run /home/cer/Desktop/cer/flinkcep/flinkcep/target/flinkicu-1.0-jar-with-dependencies.jar calhost:9092 --zookeeper.connect localhost:2181 --OUT\_GAP GAP --OUT\_COTRAVEL COTRAVEL --OUT\_COURSE COURSE***
* ***$ Kafka2.2:bin/kafka-console-consumer.sh --bootstrap-server localhost:9092 topic AIS***
* You can check the events that detected at each job at this url. http://localhost:8081/#/overview
* You can see what is send at each topic running the consumer of each topic. For example for topic AIS , running the consumer you will see the AIS messages
* ***./AIS .py AIS***
* Run the python script in order to full fill the topic that contains all the AIS messages and 1st module starts receiving AIS messages
* Producer is inside the project at the path ‘flinkcep/producer’

Running the script suspicious events will start to be detected. All of these will be written at the path *home/cer/Desktop/temp/* as .txt /.csv files. There will be a .txt /.csv file for each trajectory and suspicious event.

## 3.12. Repository

* The source code for the flink project is located at:

<https://github.com/salevizo/flinkcep.git>

## 3.13. Source code of flink project

* 2 java maven projects: implement flink patterns for detecting trajectory or complex events
  + *Flinkcep*
  + *Cep\_flinkcep*
* 1 python script *AIS .py* inside flinkcep that send AIS messages inside kafka producer
* 7 python scripts to convert the outcomes of the projects .txt files to csv in order to plot them on qgis, located inside folder */convert\_txt\_to\_csv.*
* 1 folder */outcomes* which contains all the generated outcomes of 4 different executions of the project.
* Running the project, the outcomes will be saved at the path /home/cer /Desktop/temp
* README file
* The project is located at the path:
* /home/cer/Desktop/cer\_2/flinkcep/
* Kafka is located at the path:
* /home/cer/Desktop/
* Flink is located at the path:
* /home/cer/Desktop/

## 3.14. Running environment

|  |  |
| --- | --- |
| OS | PRETTY\_NAME="Ubuntu 18.04.1 LTS"  VERSION\_ID="18.04" |
| Postgres | postgres=# SELECT version();  PostgreSQL 10.5 (Ubuntu 10.5-0ubuntu0.18.04) on x86\_64-pc-linux-gnu, compiled by gcc (Ubuntu 7.3.0-16ubuntu3) 7.3.0, 64-bit |
| QGIS | 2.18.0 |
| Python | Python 2.7.15 |
| Kafka | kafka\_2.11-1.0.0 |
| Flink | flink-1.6.2 |
| Scala | Scala code runner version 2.11.12 -- Copyright 2002-2017, LAMP/E |
| Prolog | YAP 6.2.2 (x86\_64-linux): |

### 3.14.1. Computer specifications

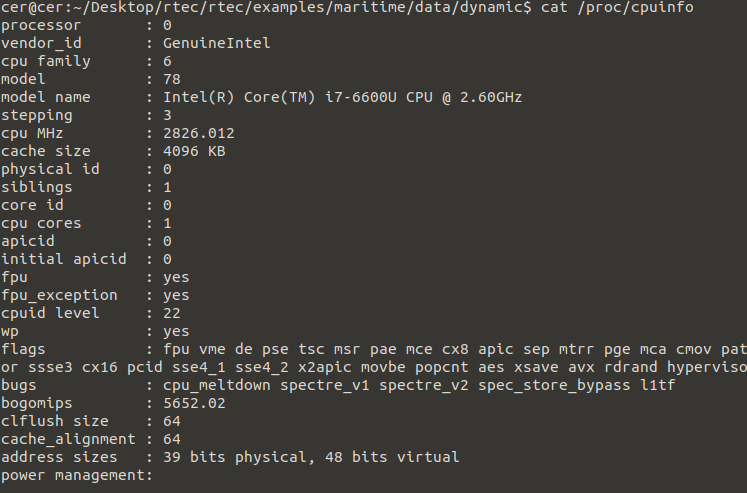


Figure 58: Cpu info

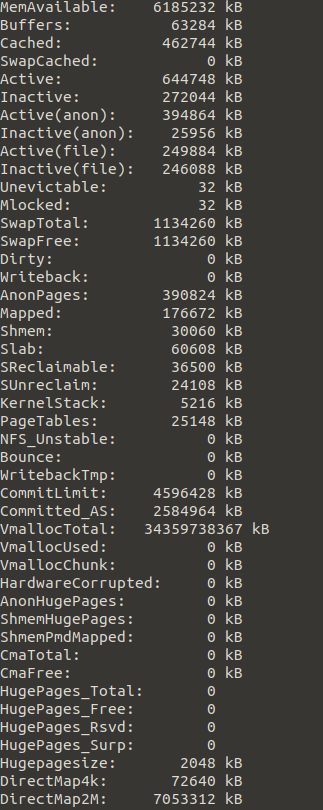


Figure 59:Memory info

# 4. Logic-based CER with RTEC

## 4.1. Complex event detecting with RTEC

Since we have already implemented a system for online monitoring marine activity over streaming positions from numerous vessels sailing at sea, at that point we will present another framework of monitoring. RTEC [3] a logic-based engine written in Prolog. More specifically RTEC it’s an event Calculus for Run-Time reasoning, designed to compute continuous queries on data streams. Event calculus represents the effects of actions on fluents. The predicate *HoldsAt* is used to tell which fluents hold at a given time point. Events are also represented as terms. The effects of events are given using the predicates *Initiates* and *Terminates*. An event won’t be terminated alone, *Terminates* predicate should be call.

RTEC accepts as input a stream of time-stamped simple, derived events (SDE)s. A SDE is the result of applying a computational derivation process to some other event, such as an event coming from a sensor, in our case the ais messages from the vessels. An advantage of RTEC system is that provides Windowing, that supports real-time query computation. The system detects abnormal behavior using the critical events that which have been detected. As critical events assumed for example, the “stop” or the “pause” of a vessel.

### 4.1.1. System Architecture

The online monitoring system uses the prolog patterns which have been compiled and an input file which contains critical events, critical events are the trajectory events we have explained on Sec. 3.2. in order to detect complex events. The output of the system is written on a txt file. There are declarations files which contain information about all the events and fluents of the scenarios we would like to detect.

There is a function ‘continuousER’ that executes all the prolog patterns on the given critical events. The time period of the events that we would like to detect should be declared on this function as long as the window size. We run the monitoring systems for window size 600secs in order to do the same executions as we have already done on flinkcep.

## 4.2. Trajectory events

We have analyzed on the previous Sections (see Sec. 3.2) what is a trajectory events. At this system, trajectory events are written on a txt file and will be given as input on the monitoring system.

A sample critical event is shown here. For example, “velocity” event means that a vessel changes its speed with a suspicious manner.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Event | timestamp\_start | timestamp\_end | mmsi | other information based on the event |
| Velocity | 1443677750 | 1443677750 | 227008170 | Xxx|xxx|xxx |

As we would like to compare the number of detected events between the two monitoring systems we have written a python script that filters the timestamp of critical events. The file final\_preprocessed\_dataset\_RTEC\_critical\_nd.csv which located at the path /home/cer/Desktop/rtec/rtec/examples/maritime/data/dynamic/ will contain all the critical events related to the time period ['1443650402','1444660250'] (same duration has been used for the ais message on flinkcep) and will be given as input on the system.

Critical events are described on the declarations files. On declaration files are declared the entities of the event description: events, simple and statically determined fluents.

Some critical events are:

* coord
* velocity
* slow\_motion\_end
* entersArea
* slow\_motion\_start
* …

## 4.3. Complex event

Complex event Recognition module consumes the input file with critical events in order to detect complex events of maritime activities. The complex event patterns are located at the path RTEC/examples/maritime/CE\ patterns/patterns.prolog. In order to run them firstly you should compile them.

The implemented patterns are listed below:

* gap
* stopped
* changingSpeed
* withinArea
* underWay
* highSpeedNearCoast
* anchored or moored
* Movement ability affected
* adrift
* rendezVous
* fishing
* lowspeed

Some of the above patterns for complex events have also been implemented and run in flinkcep so we can compare their results and highlight the differences between the 2 monitoring systems, whereas for some of them there is no special pattern on flinkcep as their functionality is included in other patterns. These patterns cannot be compared.

Looking the patterns on flink ( Java patterns) and the patterns on RTEC ( prolog patterns) we can see that the source code for RTEC patterns are less complex and more user friendly. Their format is like a pseudocode and it is easy to be maintained. Each pattern has an “Initiated” and one or more “Terminated” predicates. For example, a vessel detected as “stopped” with the following pattern:

|  |
| --- |
| InitiatedAt(stopped(Vessel)=Status,T):- |
| happensAt(coord(Vessel,Lon,Lat),T), |
| portDistance(Lon,Lat,Status). |
| terminatedAt(stopped(Vessel) = Status, T) :- |
| happensAt(stop\_end(Vessel), T). |
| terminatedAt(stopped(Vessel) = \_Status, T) :- |
| happensAt(gap\_init(Vessel), T). |

Below are described the patterns running in RTEC monitoring system. For each detected event will be presented the coordinates at which this event took place. We will use them to visualize the results on QGIS. Coordinates found after searching on psql database for the specific timestamp.

### Stopped & lowspeed

This pattern has not been implemented as special pattern on flinkcep as we check for stopped vessels inside other patterns. For example: in the Long stop pattern “stopped” is included., see Sec 3.2.5

1. WithinArea

WithinArea pattern on flinkcep is implemented via the grid partioning, see Sec 3.1.4. Here in RTEC “within area” is a pattern which detects that a vessel is inside an area for a time period. Similarly, there is no standalone pattern on flinkcep for this.

### Underway & anchored or moored

On flinkcep these patterns are included in other patterns by checking if the speed of a vessel is bigger than 5KNOTS -under way or smaller than 1 KNOT- anchored or moored. For example, see Sec 3.2.6 which detects vessels with false status. There aren’t special patterns for these.

### ChangingSpeed

ChangingSpeed pattern hasn’t been implemented as a standalone pattern on flinkcep, but inside other patterns like: HighspeednearPort, see Se 3.2.4 there is a statement which check if a vessel has changed its speed rapidly.

### HighSpeedNearCoast

The highSpeedNearCoast pattern is implemented somehow different on flinkcep. On flinkcep, we have written a pattern which detect vessels near port, more specific than detect vessels near cost.

### Gap in communication

This pattern is implemented at both systems with the same way .

### Adrift

This pattern is implemented at both systems. Nevertheless, on flinkcep we have used different approach about when a vessel seems to be in adrift, so we have changed the RTEC prolog pattern according our implementation on flinkcep in order to have a fair comparison.

### Rendezvous , Fishing

These patterns are implemented at both systems with the same way. The results of 2 systems will be compared.

## 4.4. Outcomes of the system

Running RTEC system for detecting suspicious events we can see that we don’t have exactly the same results with the previous monitoring system, Flink even though we used the same input data. It should be noticed that patterns between flinkcep and RTEC have not exactly the same context.

The below results came after running the monitoring system for time period of ['1443650402','1444660250'].

### 4.4.1. Running patterns for different window size

|  |  |  |
| --- | --- | --- |
| Detected event | 300 secs | 600 secs |
| gap | 82003 | 41508 |
| stopped | 25064 | 25299 |
| changingSpeed | 6456 | 7281 |
| atAnchorOrMoored | 3821 | 3726 |
| fishing | 2969 | 5625 |
| adrift | 2367 | 5103 |
| speedLTMin | 34589 | 21447 |
| highSpeedNearCoast | 168 | 252 |
| rendezVous | 2 | 45 |

﻿

It is obvious that with bigger window size less trajecory events are detected as the patterns become stricter, eg: gap event. On the other hand, with bigger window size the detected events come from complex events whose duration should be bigger are more. Looking for example on rendezvous complex event, using 300 secs as window size only 2 events are detected whereas when window size increased from 300 secs to 600 secs, 45 evets are detected. Meaning that inside 300 secs the “terminates” predicated hasn’t happened.

The results from several running are located at the path:/results in the folders results\_win300 and results\_win600 (window size 600 and 300 secs).

### 4.4.2. Ground truth for detected events with window size: 600 secs

We have used as window size 600 secs in order to be the same with flinkcep patterns. From the detected events on flinkcep project, we have concluded on that window size of 600 secs detects events that correspond to the ground truth.

#### ﻿ High Speed near port / coast

|  |  |
| --- | --- |
| Monitoring System | Detected events |
| FlinkCep – near port | 41 |
| RTEC – near coast | 252 |

#### Rendezvous

|  |  |
| --- | --- |
| Monitoring System | Detected events |
| FlinkCep | 58 |
| RTEC | 45 |

﻿

Comparing the outcomes of the monitoring system we see that they have some differences.

## 4.5. Visualization of the detected events

The below pictures come from QGIS framework using the results from RTEC. These visualizations represent the detected events for the patterns which describes in previous chapter, see SEC 4.3. We searched on psql database at which we have stored AIS messages and searched latitude and longitude coordinates for each timestamp in combination with vessel mmsi. We have also used as window size 600 secs in order to be the same with flinkcep patterns.

### High Speed Near Coast

The blue points are results from RTEC (252 detected events) and the red from flinkcep (41 detecetd events). As ports is a subgroup of coast, the detected events “near coast” are more than events “near port”, as it was expected.

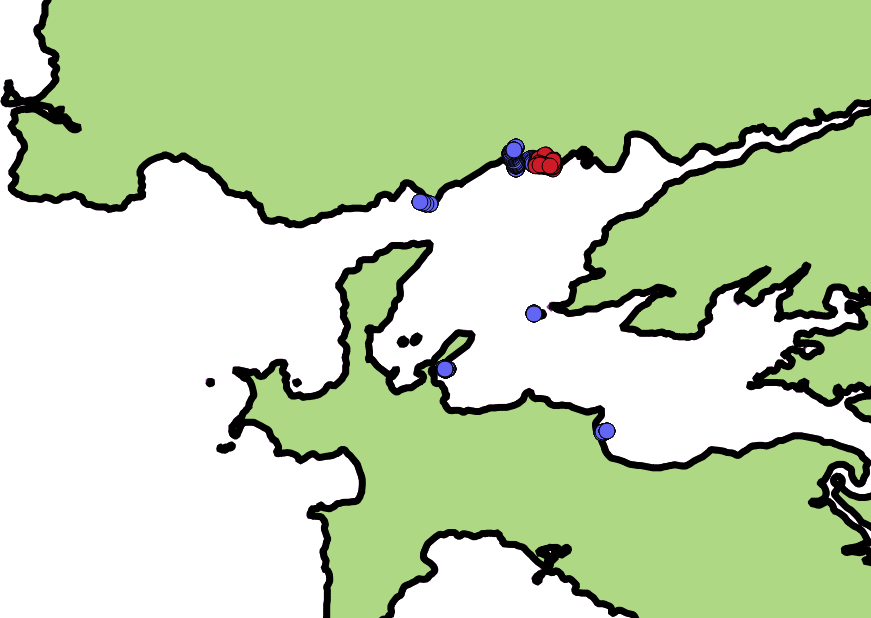


Figure 60: High speed near coast (blue: RTEC, red :flink)

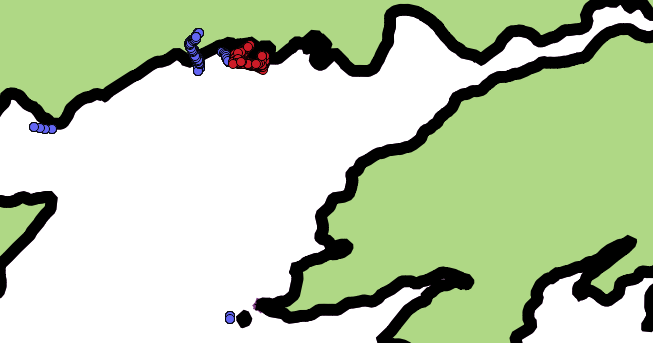


Figure 61: High speed near coast in zoom (blue: RTEC, red :flink)

### rendezVous

The blue points are results from RTEC (45 detecetd events) and the red from flinkcep (58 detected events). We compared RTEC results with the pattern running on flinkcep for index 8 (grid of *38.2m x 19m).* Flincep detects more events because of the distance rule it uses.

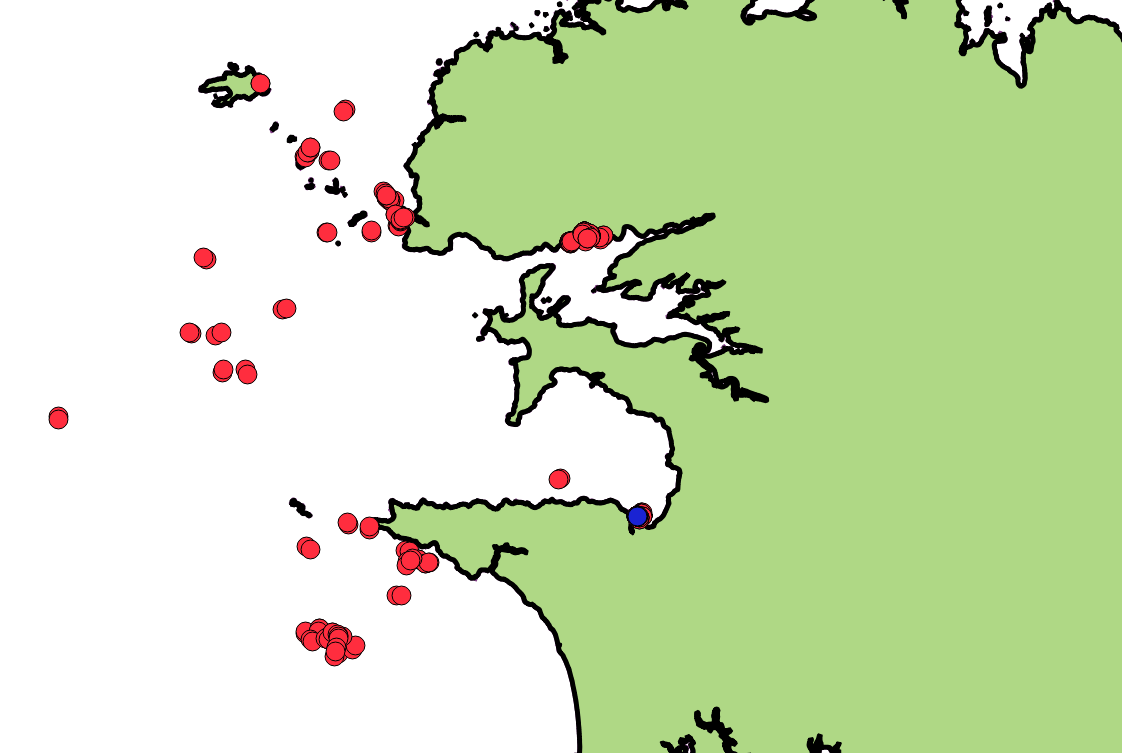


Figure 62: Vessels Rendezvous (blue: RTEC, red :flink)

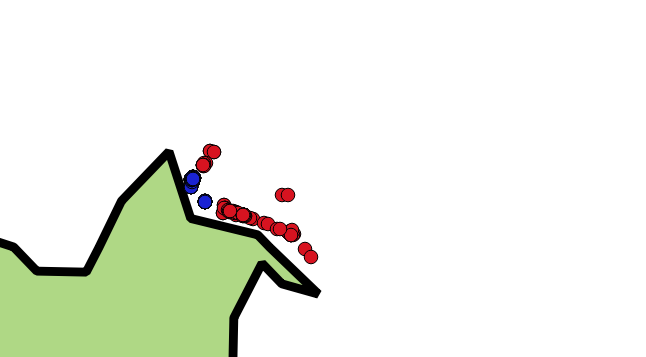


Figure 63:Vessels Rendezvous in zoom (blue: RTEC, red: flink)

### Adrift

Below is displayed a plot with all adrift detected events on RTEC monitoring system.

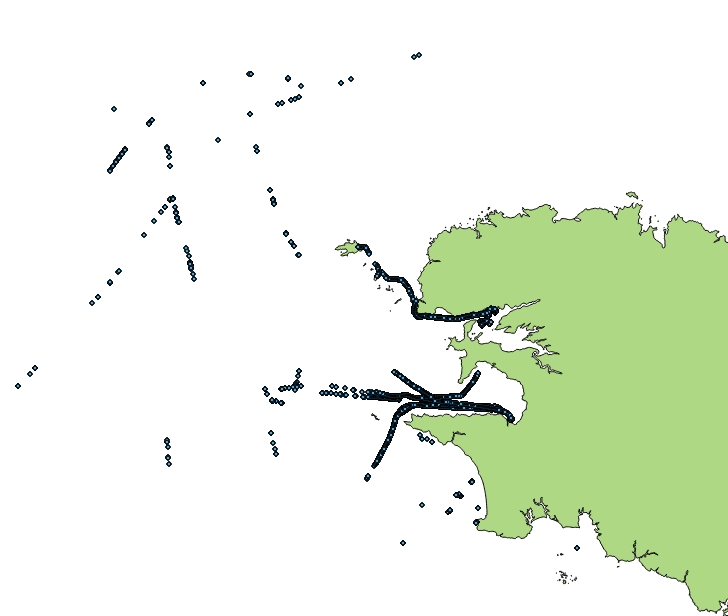


Figure 64: Adrift from RTEC

The involved mmsi at the next plot is 228064900. The blue points are results from RTEC and the red from flinkcep.

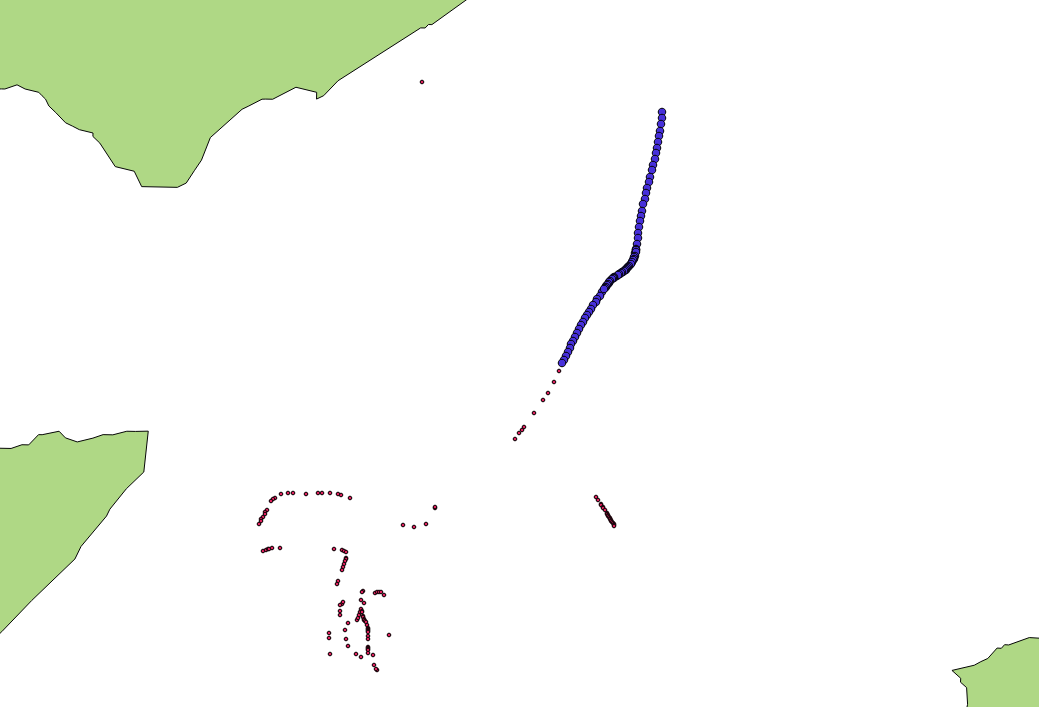


Figure 65:Adrift activity for mmsi 228064900 (blue: RTEC, red: flink)

The results from the 2 monitoring systems for the same mmsi are different enough.

## 4.6. Time of execution

The monitoring of about 1048576 critical events and the detection of complex events based on the patterns described on the Sections took less than 1 h. Faster running than flinkcep monitoring system.

## 4.7. Running commands

In order to execute the prolog patterns, they should be compiled first. Below are listed the commands you must execute.

* **Go at RTEC/examples/maritime/CE\ patterns/**
* **$yap**
* open YAP prolog
* **?- ['../../../src/RTEC.prolog'].**
* Load the main file of RTEC.
* **?-compileEventDescription('declarations.prolog','patterns.prolog','compiled\_patterns.prolog').**
* **Go at RTEC/examples/maritime/execution\ scripts/**
* **$yap**
* **?- ['continuousQueries.prolog'].**
* **?- ['../CE\ patterns/declarations.prolog'].**
* **?- ['../CE\ patterns/compiled\_patterns.prolog'].**
  + load the declarations, and the compiled patterns...
* **?- ['../data/static/loadStaticData.prolog'].**
* **?- ['../utils/loadUtils.prolog'].**
  + load static data, and some utils, both necessary for recognition.
* **%continuousER(DatasetFile,RecognisedCEsFile,RecognitionTimesFile,InputStatisticsFile,StartTime,EndTime,Step,Window)**
* continuousER('../data/dynamic/final\_preprocessed\_dataset\_RTEC\_critical\_nd.csv','../results/maritime\_results.txt','stats\_times.txt','stats\_input.txt',1443650402,1444660250,x,x).
* Call 'continuousER' with the appropriate parameters to perform recognition for a specific time period and window size.

## 4.8. Repository

* The source code for the RTEC project is located at:

<https://github.com/salevizo/rtec.git>

## 4.9. Running environment

### 4.9.1. Computer specifications

Running on the same computer which we run the flink monitoring system. For the specifications see Sec 3.14.1.

## References

[1] Online Event Recognition from Moving Vessel Trajectories Kostas Patroumpas · Elias Alevizos · Alexander Artikis · Marios Vodas · Nikos Pelekis · Yannis Theodoridis.

[2] COMPARISON OF INS HEADING AND GPS COG, R. Michael Reynolds November 6, 20

<http://www.rmrco.com/docs/m1227_Compare-cog-hdg.pdf>

[3] https://github.com/aartikis/RTEC