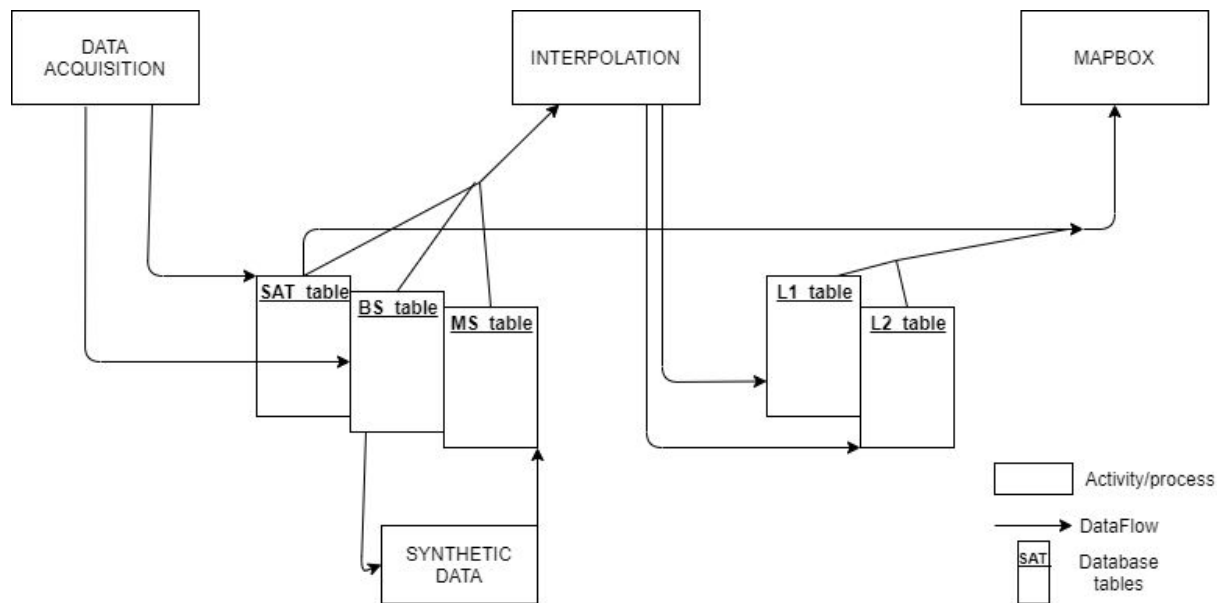# 1.GENERAL VIEW

As previous general documents, all documentation from each process is updated with separated documents.

- 4 main programs: Data acquisition, synthetic data, interpolation, mapbox
- Database divided in 2 parts; Raw data and interpolated values
- For all detailed information, see the documentation of individual programs.

2.Database Tables

Primary key: ID ascendent integer. Don't used.

Foreign keys: Non used.

BS_list

BS_list
BS_name (varchar)
Time (Datetime)
Latitude (float, as bellow)
Longitude
CO
CO2
NO2
SO2
O3
PM10
PM2,5
ID ascendent integer

SAT_list

SAT_list
Time
Latitude
Longitude
CO
CO2
NO2
SO2
O3
PM10
PM2,5
ID ascendent integer

MS_list

MS_list
Time
Latitude
Longitude
NO2
CO2
MS_indicator
ID ascendent integer

INTERPOLATED_layer1

INTERPOLATED_layer1
Time
Latitude
Longitude
CO
CO2
NO2
PM10
PM2,5
O3
SO2
GENERAL_POLLUTION
ID ascendent integer

INTERPOLATED_layer2

INTERPOLATED_layer2
Time
Latitude
Longitude
CO
CO2
NO2
PM10
PM2,5
O3
SO2
GENERAL_POLLUTION
ID ascendent integer

3. MEMORY FIRST STUDY

The maths are done by a surface of (1200km^2 aprox).

To be considered:

- From **satellite,** a total of 510.000 rows per day are stored, weighing 22MB aprox, corresponding to the surface of Portugal + Spain + France. => 154 Mb weekly. with 3.570.000 rows.
- From **Base stations,** for the same surface, we have a total of 780*24(h) = 18.720. The same occupation is assumed for any source, number of rows/columns is what matters. So with 18.720 rows we have = 0.8MB. => 5'6Mb per week.
- With the **city scooters,** from Reby we have 50.000 rows per week, equivalent to 2'15Mb. Measures are taken with a periodicity of 12 seconds.But when maths are done: ((50.000/7)/24)/300, we noticed something wrong because it results a negative number of sensors. So, we supposed a **fleet of 100 scooters,** giving us values every 10 minutes: 6*100*24*7= 100.800 rows, traduced into = 4'34Mb.

From the interpolated part:

- For **layer 3**(street level), with a Barcelona-like surface we store 1000 points per 30 minutes, traduced to **weekly numbers** ; 336.000 rows = 14'5Mb.
- With **layer 2** we use 4500 points per hour to represent catalonia's surface;

  756.000 = 32'6Mb weekly.

Extrapolations:

- We suppose that portugal + spain + france have 10 "Barcelonas" cities. Giving a total from: 14'5*10= 145Mb weekly
- Catalonia's surface is aprox, 1/30 from total source taken into account, then: 32'6*30 = 978Mb.

Reduction of memory strategies:

1. Layer 1 (Satellite) is taken directly from database, without interpolated values, so we don't need to store a Layer 1 table.
2. Interpolated values will be just stored for short period of time, for example 1 week. So when a user needs data from past 7 days, mapbox does a request to interpolation program and process the data. By using that, we just need to store the historical values ones.

Available memory:

With less than 1000 euros, we can get 16TeraBytes of memory, reducing for Raid purposes, 14Tera lasts.

Total weeks with 14 tera:

"10 Barcelonas"

- D = Data from source = 154 + 10*5'6 + 10*4'34 = 253'3 Mb.
- N = Number of weeks
- I = Interpolated values (week memory, just 1 time to be considered) = 978 + 145 = 1123Mb.
- T = Tera
- $D * N + I = 14 * T \Rightarrow N = \frac{14*T-I}{D} = \frac{14.000.000 - 1123}{253'3} = 55.331$ weeks aprox; = 55.331/52 = **1.000 years. !!?**

Knowing that we have to extrapolate it around the world, those numbers seems suspicious for me, i don't think if it is well done, we have to recheck. Considering maths well done, at least for a selected european environment, it not seems to be a matter. By the time memory will be over, there will probably exist better storage systems.

SYSTEM DESIGN (comments from Carles)

- When is data processed?
  - The first processing it's done on data acquisition program, we convert concentration and AQI values, from satellites and base stations respectively, to CAQI ones, as the European legislation.
  - Secondly, interpolation process uses the 3 data source tables, fusion them depending on which layer is making and store the processed data.
- Where is the output stored?
  - Output from interpolation is stored in 2 tables in database, then mapbox uses them and
- How are the different zoom levels handled?
  - Starting by the "users view", mapbox algorithm is able to detect the current zoom looking at, so depending if you pass through a threshold, the picture seen changes. Interpolation gives a list of points, belonging to an equidistributed grid, each value is merged by using 1, 2 or the 3 types of source, according in which layer to draw.
- Sensor data fusion, mobile sensor calibration, visual interpolation.

Starting from data mobile sensor calibration, at the moment, we don't find an ideal solution for it. We know that Reby uses 3 sensors at the same scooter, we want to store just 1 in database, so a solution could be an average between them. Visual interpolation is done via mapbox, given a grid with values from interpolation. The accuracy of the greed depends on the zoom.