



Treball Final de Màster

Estudi: Màster en Enginyeria Informàtica

Títol: Recopilació i anàlisi de les dades obtingudes d'una xarxa de sensors per analitzar la qualitat de l'aire d'una ciutat

Document: Resum

Alumne: Robert Garcia Ventura

Tutor: Dr. Santiago Thió Fernández de Henestrosa

Departament: Informàtica, matemàtica aplicada i estadística

Àrea: Estadística i investigació operativa

Convocatòria (mes/any): Setembre/2019

Actualment cada vegada es pren més atenció a la salut de les persones. En concret, la contaminació de l'aire ha esdevingut un problema molt important, especialment a les grans ciutats, on la gran quantitat de vehicles i indústries provoca una gran contaminació de l'aire. Segons la Unió Europea diversos països van superar un o diversos dels seus límits d'emissió relatius a quatre importants contaminants atmosfèrics al 2010. A finals de juliol del 2019, la Comissió Europea va denunciar a Espanya perquè des del 2010 a les ciutats de Barcelona i Madrid s'havien superat els límits de qualitat de l'aire establerts per la Comissió Europea.

El projecte RoMi neix inicialment de dos altres projectes realitzats amb la col·laboració de l'estudiant Miquel Farreras Casamort durant el primer any del Màster en Enginyeria Informàtica a la Universitat de Girona. Posteriorment, aquest projecte, gràcies al professor Dr. Nicolas Boccard Marie Maurice, es va continuar a través d'un conveni de col·laboració amb la Universitat de Girona.

Al poc de començar el projecte, es va presentar la idea al concurs SmartCAT Challenge 2019, que organitza anualment la Generalitat de Catalunya, i vam guanyar el 3^r premi.

A l'inici, el projecte es va dividir en dos parts. Per una banda, el disseny de nous prototips d'estacions de mesurament de la qualitat de l'aire amb més sensors, i per l'altre banda, la recopilació i anàlisi de les dades provinents de la xarxa de sensors.

L'estudiant Miquel Farreras Casamort es va encarregar del disseny i el muntatge de les estacions en el seu treball final de màster.

En aquest projecte s'ha realitzat el disseny del sistema de recopilació de dades altament escalable, així com la complementació del projecte amb un anàlisi de les dades de la qualitat de l'aire de Catalunya.

El disseny del sistema de recopilació de dades s'ha basat, des del principi, en dos punts molt importants:

- **Escalabilitat:** El sistema de recopilació de dades havia de ser altament escalable i estar preparat per rebre dades provinents de milers d'estacions.

- **Cost:** El sistema hauria de ser el màxim d'econòmic possible per poder-lo mantenir en funcionament a llarg termini

El disseny i creació del sistema de recopilació de dades (SRD) s'ha realitzat en 4 etapes, on a cada una d'elles s'han anat millorant i corregint els problemes trobats.

En la primera etapa es va configurar un servidor privat virtual (VPS) amb el sistema operatiu Ubuntu amb Plesk. En al mateix servidor es va configurar una base de dades MySQL amb una única taula per emmagatzemar les dades recollides per les estacions del projecte RoMi. Aquest servidor només disposava de 1 GB de memòria RAM i 1 vCPU. Això va esdevenir un problema quan es van afegir més estacions, ja que la base de dades es va col·lapsar per falta de memòria i potència de càlcul.

En la segona etapa es va desacoblar la base de dades del servidor i llavors es va crear una base de dades serverless utilitzant el servei Amazon RDS que ofereix l'empresa Amazon Web Services (AWS). Amb aquest servei es va aconseguir separar el rendiment del servidor i la base de dades. Amb aquest canvi, es va millorar molt el rendiment del servidor, i les peticions de les estacions es van poder processar sense cap problema.

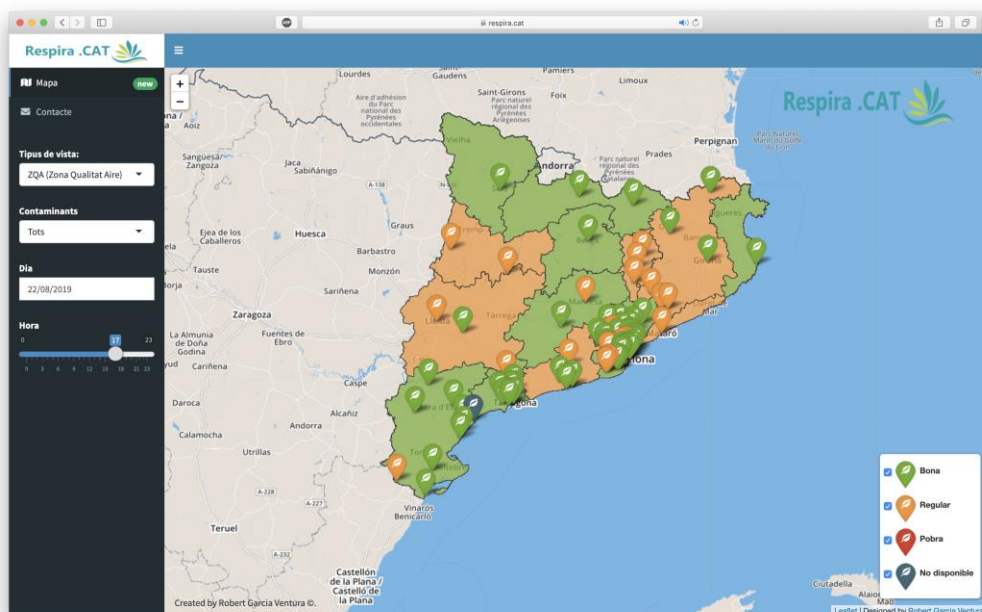
En la tercera etapa es va eliminar el VPS per les dificultats que afegia la seva escalabilitat i els costos que hauria suposat de cares al futur. Com a solució alternativa al VPS es van utilitzar dos serveis de AWS completament serverless, altament escalables i amb un cost molt reduït. Per a l'enviament de les dades es va dissenyar una API REST i es van implementar tots els endpoints de la API amb el servei AWS API Gateway. El backend de tots els endpoints de la API es van programar utilitzant el servei AWS Lambda. Amb aquesta solució, a part d'haver millorat molt la comunicació de dades entre la base de dades i les estacions gràcies a l'API, també va permetre aconseguir un sistema altament escalable amb un cost molt reduït.

En la quarta i última etapa es va millorar la seguretat de la comunicació de les dades amb l'API, afegint un token de seguretat utilitzant el servei AWS Cognito. També es van afegir dades de qualitat de l'aire i meteorològiques de fonts externes per utilitzar-les posteriorment en l'anàlisi de dades. Per saber l'estat de les estacions del projecte RoMi es va crear un bot de Telegram i programa't una tasca diària amb Amazon CloudWatch per saber l'estat de les estacions a l'inici de cada dia.

Amb les dades provinents de les estacions del projecte RoMi es va crear la pàgina web www.romibox.com per visualitzar en temps real la contaminació de l'aire de les estacions del projecte RoMi.

Amb les dades provinents de les estacions de la Generalitat de Catalunya es va crear la pàgina web www.respira.cat, per visualitzar en temps real i l'històric de la contaminació atmosfèrica a tot Catalunya.

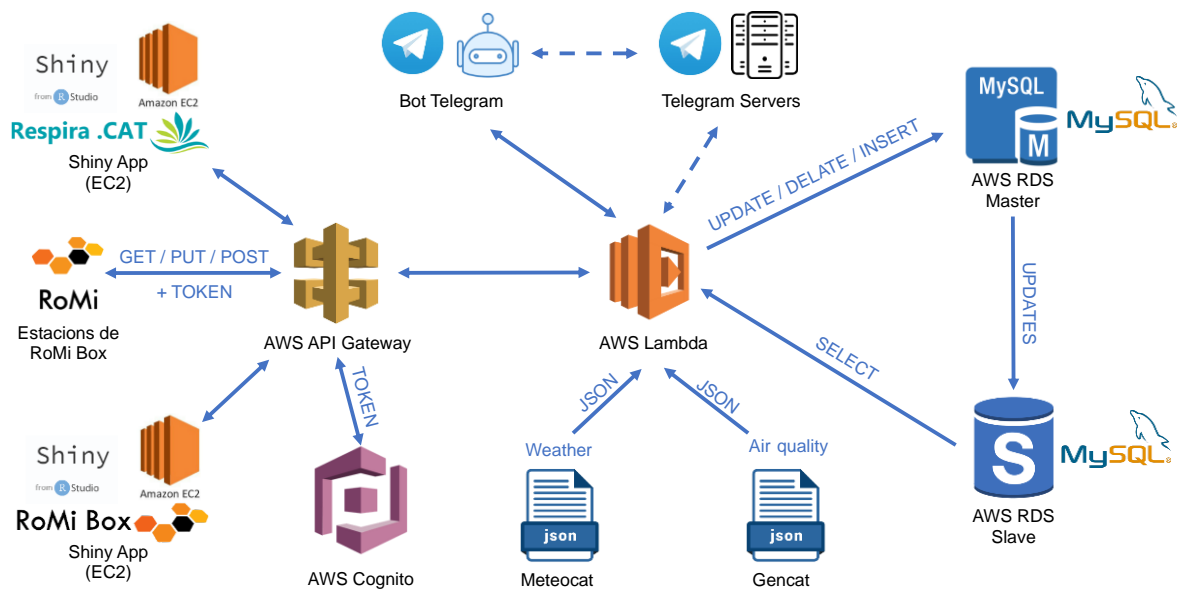
A la Imatge 1 es pot veure la pàgina web de Respira .CAT amb la qualitat de l'aire agregada per zona de qualitat de l'aire (ZQA).



Imatge 1 Pàgina web de Respira .CAT agrupant les dades per zona de qualitat de l'aire (ZQA)

Al crear les dos pàgines webs i augmentar el nombre d'estacions del projecte RoMi, la quantitat de consultes a la base de dades es va incrementar de forma considerable. Per millorar el rendiment de la base de dades, es va crear una segona base de dades replica (esclau) de la base de dades principal (mestre). D'aquesta forma, totes les peticions d'escriptura es realitzen a la base de dades principal i totes les peticions de lectura a la base de dades replica.

En la Imatge 2 es pot veure de forma visual l'esquema del sistema de recopilació de dades creat a la quarta etapa.



Imatge 2 Disseny de la quarta etapa del sistema de recopilació de dades

A l'anàlisi de la qualitat de l'aire s'han analitzat els diferents contaminants i el seu efecte amb les dades meteorològiques. S'han definit com a variables objectius el PM10 i O₃ (ozó), i s'ha predit la qualitat de l'aire amb dos models.

En la primera part de l'anàlisi s'han obtingut les dades de la base de dades, s'han filtrat i aplicat transformacions de les variables per posar-les en el format correcte.

Posteriorment, s'han imputat els valors de la qualitat de l'aire de les estacions que els hi faltaven valors. Per imputar les dades, s'ha calculat les estacions més pròximes per a cada una de les estacions i, posteriorment, per a cada valor que s'hauria d'imputar, s'ha obtingut el valor de l'estació més pròxima amb el valor existent.

Un cop s'han imputat totes les dades de la qualitat de l'aire, s'han correlacionat amb les dades meteorològiques per posteriorment crear els models.

Finalment, en l'última part s'han creat dos models ANCOVA i dos series temporals ARIMA per poder predir la qualitat de l'aire de Catalunya a curt termini a nivell d'estació de qualitat de l'aire.