



Universitat
de les Illes Balears

TRABAJO DE FIN DE MÁSTER

DESPLIEGUE DE UNA INFRAESTRUCTURA DE ANÁLISIS DE DATOS PARA LA GESTIÓN DE LA CALIDAD DEL AIRE EN UN CENTRO ESCOLAR

Arantxa Oteiza Betelu

Màster Universitari en Anàlisi de Dades Massives en Economia i Empresa (MADM)

Centro de Estudios de Postgrado

Año Académico 2020-21

DESPLIEGUE DE UNA INFRAESTRUCTURA DE ANÁLISIS DE DATOS PARA LA GESTIÓN DE LA CALIDAD DEL AIRE EN UN CENTRO ESCOLAR

Arantxa Oteiza Betelu

Trabajo de Fin de Máster

Centro de Estudios de Postgrado

Universidad de las Illes Balears

Año Académico 2020-21

Palabras clave del trabajo: Sensores, Calidad del aire, Infraestructura, Análisis, Big Data, Salud

Nombre Tutor/Tutora del Trabajo: Isaac Lera Castro

Resumen

En la actualidad, afrontar la analítica de datos como un paso necesario para avanzar en la digitalización, es prácticamente de carácter obligatorio. Aun así, existe una gran cantidad de entidades que siguen prescindiendo de la generación de beneficio en base a la información.

Un caso específico es el de las instituciones educativas, las cuales suelen carecer de herramientas para poder hacer uso de las ventajas y oportunidades que ofrece el mundo del *Big Data*¹.

Este proyecto consiste en la creación de una infraestructura y arquitectura de datos para el Colegio Alemán de Zaragoza.

La preocupación relativa al estado de salud de los alumnos frente a la pandemia del COVID-19, ha provocado la necesidad inmediata de contar con un sistema de análisis de la calidad del aire en las aulas.

A lo largo del proyecto se irá demostrando la importancia de una buena estructura de software y hardware, así como el posterior análisis y desarrollo de los datos obtenidos mediante diferentes técnicas y herramientas.

Abstract

Nowadays, addressing data analytics problem is practically a mandatory step to achieve the best practices for digitalization success process. Even so, there are a large number of entities that do not take advantage of the benefits based on information. A specific case are the educational institutions, which often lack the tools to make use of the advantages and opportunities offered by the world of Big Data.

This project consists of the creation of a data infrastructure and architecture for the German School of Zaragoza. The concern about the health condition of students in terms of the COVID-19 pandemic, has led to the immediate need for a system to analyse the air quality in their classrooms.

Throughout the project, we analyse the importance of a good software and hardware infrastructure, as well as the subsequent analysis and development of the data obtained through different techniques and tools, for achieving a successful data analytics implementation.

1. Introducción

En un mundo cambiante y dinámico como el actual, donde constantemente tienen que tomarse decisiones y medidas que pueden cambiar el devenir de una empresa, institución, negocio e incluso, de nosotros mismos, la transformación digital es indispensable para la supervivencia.

Dentro de un entorno tan globalizado en el cual se consiguen grandes volúmenes de datos de forma constante, existe una necesidad de gestionar, almacenar, procesar y analizar con la finalidad de extraer información relevante sobre los datos generados.

La crisis generada por la pandemia del COVID-19 ha provocado la necesidad inmediata de llevar a cabo un proceso de digitalización en aquellos entornos más obsoletos [1]. Desde el inicio de la pandemia, el no contar con los medios adecuados ha supuesto poner en riesgo las vidas de las personas. Esto ha obligado al mundo a poner en valor la importancia del buen uso y control de datos.

El impacto de la crisis sanitaria ha sido de gran envergadura en cualquier ámbito, pero uno de los que menos preparado estaba, era el educativo. Fue por ello por lo que desde el gobierno se exigía una readaptación de las infraestructuras para poder seguir adelante con la enseñanza en las aulas [2].

Es común que los colegios e instituciones dedicadas a la enseñanza estén dotadas de una tecnología básica con la cual los alumnos puedan aprender, pero, en su mayoría, no con un desarrollo suficiente como para poder hablar de sistemas totalmente digitalizados.

La situación crítica que se ha vivido ha destapado una carencia en los colegios [3]. Ésta es la falta de una infraestructura sólida en la que se puedan apoyar para tomar decisiones inteligentes basadas en datos reales y objetivos.

Como consecuencia a estas carencias y necesidades descubiertas, nace este proyecto de digitalización del Colegio Alemán. El objetivo de este proyecto es la creación de una infraestructura que permita a la

¹<https://www.ibm.com/analytics/hadoop/big-data-analytics>

escuela tener un sistema de generación, almacenamiento y análisis de datos sencillo que los lleve a superar con éxito la etapa de escolarización durante la pandemia contando con el requisito de educación presencial.

El proyecto descrito en este documento surge a través de una contratación de la entidad educativa a la empresa Belerofontech² en la cual soy analista de datos. El colegio acudió en solicitud de asesoramiento dada su situación de inexperiencia en el sector y necesidad de renovación e implantación de nuevas tecnologías basadas en datos.

El objetivo principal de este trabajo es proporcionar al Colegio Alemán de Zaragoza³ una infraestructura de captura y análisis de datos totalmente accesible por cualquier persona del centro que permita sentar las bases tecnológicas para futuros proyectos.

Sujeto al objetivo principal, vienen cuatro requerimientos que se deben ir cumpliendo para considerar exitoso el proyecto. Estos cuatro requerimientos se dividen a su vez en dos temas principales: infraestructura y análisis.

Los requerimientos relacionados con la infraestructura son:

- *Monitorizar los valores ambientales proporcionados por los dispositivos instalados en las aulas con la finalidad de tomar medidas y mejorar la calidad medioambiental en las aulas.*

Se pretende crear en el colegio una red de dispositivos mediante los cuales se puedan medir indicadores ambientales que demuestren la calidad del aire en cada aula.

- *Proveer de manera pública de la visualización de los datos obtenidos, así como sus análisis.*

No se trata de un proyecto que pretenda hacerse única y exclusivamente para este colegio. La idea es que, si algún otro centro o institución puede aprovechar los datos o la metodología aplicada para la creación de la infraestructura, pueda hacerlo basándose en este ejemplo.

Por otro lado, los requerimientos relacionados con el análisis son:

- *Analizar los datos extraídos de los sensores y tomar decisiones en base a ellos.*

Se pretende que la plataforma facilite la toma de decisiones sobre el estado ambiental actual del aula a profesores como alumnos.

- *Extraer conclusiones firmes en base a los datos obtenidos.*

Al contar con una gran cantidad de datos debido al volumen de aulas, se busca encontrar patrones o diferencias entre ellas con la finalidad de alcanzar el confort ambiental en todas ellas por igual.

Dentro de este requerimiento se pretenden resolver dos cuestiones más específicas: ¿Existe relación entre el CO2 en las aulas y las temperaturas? ¿Existen diferencias de confort térmico en las aulas?

Todos estos requerimientos abordados a lo largo del trabajo hacen necesaria una inversión en software y hardware para poder ser llevados a cabo. Es por ello por lo que, a continuación, se describe la arquitectura que se ha decidido utilizar en el proyecto.

2. Arquitectura

Para poder desarrollar el proyecto, es preciso definir todos aquellos elementos que conforman la estructura final del mismo.

Uno de los puntos más importantes del trabajo es la captación de datos mediante sistemas de medición de calidad del aire precisos. Para conseguir esta captación de datos, un requisito indispensable es hacerse con el hardware necesario e instalarlo en el centro.

Una vez obtenido, se comienza con la instalación y configuración de diferentes softwares para llevar a cabo todos los pasos planteados.

2.1. Elementos hardware

El proyecto tiene como objetivo dispositivos hardware de coste ajustado, con arquitecturas eficaces para su cometido y con una amplia comunidad de usuarios que permita obtener soporte y apoyo.

Las aulas deben contar con pequeñas instalaciones capaces de integrar un número concreto de sensores y, a su vez, tienen que ser capaces de transmitir esta información a colectores de datos más centralizados que almacenan y tratan dicha información. Los elementos utilizados para tal fin son los Arduinos y los Sensores Grove.

²<http://www.belerofontech.com/>

³<https://colegioaleman.com/>

Un Arduino⁴ es una plataforma electrónica de código abierto basada en hardware y software libre. Por ejemplo, las placas Arduino son capaces de leer una entrada proporcionada por un sensor de luz, por la pulsación de un botón, o incluso la detección de un mensaje de Twitter. Convirtiendo esta entrada en otra salida lógica que puede, por ejemplo, activar un motor, encender un emisor de luz, o publicar algún mensaje en alguna red social de Internet.

A lo largo de los años, Arduino ha sido el cerebro de miles de proyectos, desde objetos cotidianos hasta complejos instrumentos científicos [4]. Una comunidad mundial de creadores -estudiantes, aficionados, programadores y profesionales se han reunido en torno a esta plataforma de código abierto y sus contribuciones han sumado una gran cantidad de conocimiento accesible que pueden ser de gran ayuda tanto para los principiantes como para los expertos.

En este proyecto, en cada Arduino de una estación de medición, se instala un programa que permite la lectura de los datos generados por todos los sensores con una frecuencia preestablecida. Esta frecuencia es configurable.

De los Arduinos podemos extraer diferentes mediciones como: movimiento, presión, sonido, humedad, partículas en suspensión, CO2 o temperatura.

A parte de los Arduinos, se han adquirido otros sensores con la finalidad de poder comparar la fiabilidad de los datos e incluso agregar información sobre la calidad del aire en el espacio estudiado.



Ilustración 1 Arduinos instalados en las aulas

Grove⁵ es un sistema modular de sensores con conectores estandarizados. La plataforma tiene un

⁴<https://www.arduino.cc/>

⁵<https://www.seeedstudio.com/category/Grove-c-1003.html>

enfoque orientado al ensamblado de componentes electrónicos, opuesto al enfoque tradicional basado en puentes o soldaduras, por lo que es más fácil de conectar, experimentar y construir y simplifica el proceso de aprendizaje.

En cada aula hay instalada una estación que mide los siguientes valores ambientales de las aulas:

- Temperatura
- Humedad
- CO2
- TVOC
- Calidad general del aire
- Concentración de PM1.0
- Concentración de PM2.5
- Concentración de PM10
- Ruido ambiental
- Iluminancia

Para poder completar la definición del hardware utilizado en el proyecto, es preciso mencionar anteriormente dos elementos software con el fin de comprender mejor la función computacional de la Raspberry PI -otro tipo de dispositivo pequeño que se describe a continuación de estos dispositivos-. Estos elementos software son NodeRed y Mosquitto.

NodeRed se utiliza en este proyecto con la finalidad de controlar el flujo de datos antes de que entren a la base de datos, ya que, según comenta Enrique Crespo en su web *Aprendiendo Arduino* [5], se ha convertido en poco tiempo en el estándar open-source para la gestión y procesado de datos en tiempo real.

También permite comunicarse con los Arduinos instalados. Por ejemplo, puede solicitarles reducir la frecuencia de almacenamiento de datos durante la noche o el fin de semana. Además, NodeRed permite el reenvío de datos, lo que nos ayuda a enviar directamente los datos a una instancia remota de PostgreSQL.

Se trata de una herramienta de programación para conectar dispositivos de hardware, APIs y servicios en línea de forma sencilla, sin necesidad de programar código.

Proporciona un editor basado en un navegador que permite la conexión de los flujos utilizando la amplia gama de tipos de nodos que pueden desplegarse en tiempo de ejecución con un solo clic.

Por otro lado, Mosquitto es un servidor de mensajes de código abierto que permite gestionar la recepción de datos provenientes de decenas de estaciones de captura de datos, en nuestro caso, los obtenidos a través de los sensores Grove y los instalados sobre los Arduinos. Estos datos son accedidos a través del software de NodeRed.

Otro elemento fundamental en la infraestructura fue la Raspberry Pi. La función de este dispositivo es principalmente concentrar los datos, recibiendo los datos capturados por Mosquitto a través del software de NodeRed.

Raspberry fue diseñada y creada originalmente por una organización sin ánimo de lucro en el Reino Unido [6], con la intención de crear una placa base para promover la enseñanza de la informática básica en las escuelas y en los países en desarrollo. Es un ordenador del tamaño de una tarjeta de crédito. A pesar de su pequeño tamaño, puede realizar la mayoría de las tareas que un ordenador de escritorio convencional. De este elemento fueron adquiridas dos unidades que se encontraban conectadas mediante USB a la red con la finalidad de controlar y formar parte del proceso de captura de datos

Para conectar todos estos elementos, el colegio ha renovado su infraestructura de red, instalando una red de fibra óptica que cubre todas las zonas de uso. Cada espacio clave del centro cuenta con una *OpticalNode Terminal*⁶ capaz de ofrecer múltiples servicios digitales. Se ha dimensionado también la infraestructura de enrutamiento, corta fuegos y mejorado su ancho de banda de salida para dar al centro los recursos necesarios para afrontar éste y futuros proyectos de transformación digital.

2.2. Aplicativo

Cualquier proyecto de IoT es intensivo en el uso de diversos elementos software que permitan articular la solución. Nuestro proyecto incorpora este tipo de elementos con el objetivo de componer una solución completa, estable y escalable.

Un elemento clave del proyecto son dos servidores, uno local y otro en la Nube. Ambas máquinas

hacen uso de sistema operativo Linux⁷, lo que nos permite emplear elementos software de licencia libre que generan la infraestructura necesaria para poder capturar y procesar los datos del proyecto.

Por otro lado, también es clave la base de datos utilizada para el proyecto, la cual se usa para el almacenamiento de los registros captados durante el proceso. La opción elegida es PostgreSQL⁸. Se trata de un sistema de base de datos relacional de código abierto con más de 30 años de desarrollo activo que le ha ganado una fuerte reputación por su fiabilidad, robustez y rendimiento [7]. El objetivo es que esta plataforma sea la base del almacenamiento de la información cuantitativa del colegio y que a medio plazo permita explotar la información generada por la comunidad educativa.

Además, en este proyecto se generará código fuente en Python y SQL, además de los proyectos NodeRed y ficheros de configuración de todas las aplicaciones.

Con la finalidad de dinamizar los procesos mencionados, se hará uso de Apache Airflow⁹ para implementar un orquestador que permita automatizar todo tipo de tareas de gestión de la información dentro del proyecto. Se entiende por orquestador a un software que permite configurar, gestionar y coordinar sistemas informáticos de manera automatizada [8].

Para el almacenamiento de todos los archivos, scripts, dashboards y, en general, diferentes archivos del proyecto, se eligió *GitHub*¹⁰.

Esta plataforma permite que todos los archivos que se creen durante el desarrollo del proyecto se puedan compartir de forma libre y segura.

Se va a crear un repositorio propio del Colegio Alemán que, en un principio, se utilizará para este desarrollo, pero la idea sería que de cara al futuro pueda utilizarse para almacenar el contenido del propio colegio.

⁶<https://open.spotify.com/>

⁷<https://www.linux.org/>

⁸<https://www.postgresql.org/>

⁹<https://airflow.apache.org/>

¹⁰<https://github.com/>

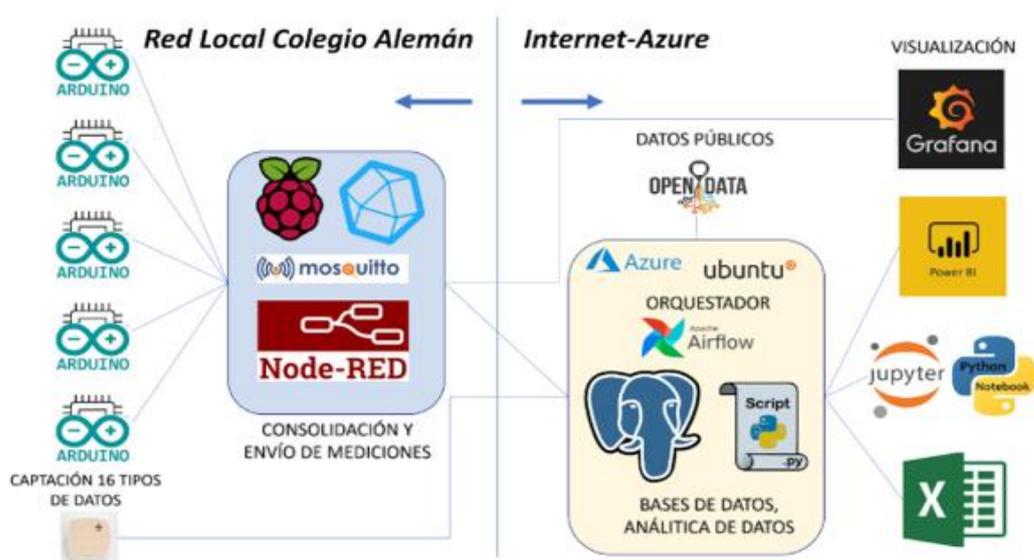


Figura 2 Arquitectura del proyecto

Finalmente, se consideran de vital importancia en el proyecto las plataformas de visualización y análisis de datos. Por ello, para llevar a cabo la parte final del proyecto, se van a utilizar herramientas accesibles como Power BI¹¹, Excel¹², Jupyter Notebook¹³ o Grafana¹⁴.

2.3. Arquitectura final

Una vez descritos todos los elementos que participan en el desarrollo final de la infraestructura, se procede a presentar la arquitectura final del proyecto. Ésta se sintetiza en la Figura 2.

Como puede verse en la Figura 2, la idea es que los Arduinos (elementos ubicados en la parte izquierda de la imagen), se comunican por el protocolo MQTT mediante wifi, estándar dentro del IoT. Estos enviarán la información capturada al servidor Raspberry Pi y a un proveedor de servicios donde reside la base de datos, en este caso Azure, haciendo uso de MQTT.

Mediante el orquestador Airflow, se consigue una consulta definida en un script de Python con un intervalo de unos 15 minutos que recoge los datos obtenidos en ese tiempo y los lleva a la base de datos PostgreSQL que reside en un servidor de Azure. Estos elementos se encuentran situados en la mitad de la imagen ya que son los intermediarios entre la captura de datos y su visualización.

Una vez llegado a este punto, se puede acceder a la base de datos desde múltiples plataformas con la finalidad de procesar, analizar y visualizar los datos.

Las herramientas elegidas para la visualización, como se ha mencionado con anterioridad, son: Excel, Power Bi, Grafana y Jupyter Notebook (situados al margen derecho de la Figura 2).

Con el objetivo de conseguir un análisis más realista y comparativo, también se acude a fuentes de datos abiertos (como en este caso, AEMET), también recogidas gracias a un script Python ejecutado a intervalos planificados en Airflow.

¹¹<https://powerbi.microsoft.com/es-es/>

¹²<https://itconnect.uw.edu/learn/workshops/online-tutorials/microsoft-office-2010/microsoft-excel-2010/>

¹³ <https://jupyter.org/>

¹⁴ <https://grafana.com/>

3. Implantación

En este apartado se va a explicar de qué manera se ha buscado conseguir los objetivos planteados en el proyecto.

3.1. Instalación de dispositivos

En cada aula se ha procedido con la instalación de un sensor Grove y en una de cada dos aulas se ha instalado también un Arduino. Esto ha sido debido a que los Arduino han servido para corroborar que los datos de los sensores Grove son realistas y correctos, por ello no es necesario instalarlo en todas las aulas, además de suponer un desembolso económico mayor.

En total, han sido instalados 40 sensores Grove y 18 Arduinos.

La elección del emplazamiento de los dispositivos en las aulas ha sido ágil por el reducido tamaño de los dispositivos. El lugar seleccionado deba cumplir con el requisito de no interferir en la toma de mediciones, es decir, contar con un ambiente neutro que no proporcione “ruido” o sesgo en la captación de ningún parámetro.

El lugar decidido ha sido la parte superior de un mueble de 170cm de alto con el que cuentan todas las aulas que se encuentra situado nada más entrar a la clase. De esta manera los alumnos no pueden acceder al dispositivo y, a su vez, se encuentra lejos de las ventanas para evitar el sesgo informativo.

3.2. Almacenamiento de datos

Para el almacenamiento de los datos capturados, se ha elegido una PostgreSQL a la que puede accederse desde cualquier cliente compatible.

Gracias a un script ejecutado a intervalos planificados con Airflow, la base de datos va recibiendo cada 15 minutos todos los registros que hayan tomado los sensores. La elección de este intervalo de tiempo viene del interés en obtener información de la calidad del aire con un grano un poco más profundo que el horario, pero no tan desglosado como para ser analizado por minutos. De ahí la inclinación a tomar cuatro registros por aula y hora.

De esta manera, a fecha de realización de esta documentación se ha obtenido más de dos millones de registros que corresponde a un periodo desde el día

cinco de enero al quince de julio de este mismo año 2021.

El lenguaje para la realización de consultas a la base de datos es SQL, puesto que se trata de una base de datos relacional

4. Análisis de datos

A continuación, se va a exponer todo el análisis realizado con la finalidad de cumplir con todos los requerimientos propuestos en la introducción del proyecto. Como se va a ver más adelante en el transcurso del análisis, varios requerimientos conllevan diferentes hitos necesarios para poder darse por cumplidos. Dichos hitos, así como su cumplimiento, se definen también en este apartado.

4.1. Requerimientos de infraestructura

4.1.1. Base de datos

A medida que se ha ido exponiendo la infraestructura del proyecto, se ha descrito la resolución del primer requerimiento. El hito que debe conseguirse para la monitorización de los valores ambientales en las aulas es la propia infraestructura en sí.

Una vez conseguida una base de datos que se va autocompletando y genera contenido gracias a los dispositivos que le envían la información ya se considera que existe una plataforma sólida mediante la cual se pueda medir la calidad ambiental en las aulas e ir más allá en cuanto a análisis y extracción de conclusiones.

4.1.2. Código compartido

El segundo requerimiento se cumple con la creación de una cuenta de GitHub en la cual el colegio ha ido publicando todo el código y material necesario para difundir públicamente los datos y análisis generados durante el proyecto.

Lo que se pretende mediante este hito es dar ideas a otras instituciones para que puedan llevar a cabo sistemas parecidos que ayuden a garantizar el bienestar de los alumnos.

Al tratarse de una plataforma gratuita, se garantiza el cómodo y asequible acceso a las visualizaciones y resultados obtenidos.

4.2. Hitos del Análisis

Una vez instalada la red de captura de datos e implantada la base de datos PostgreSQL como se ha definido anteriormente, se aborda el segundo objetivo definido: El análisis y toma de decisiones en función de los datos.

Para el cumplimiento de este objetivo se han establecido los siguientes hitos:

H1. Conocer en qué momento los niveles de CO2 superan los niveles permitidos y llevar a cabo acciones de aviso.

Para cumplirlo, es necesario llevar un control sobre los registros que llegan a la base de datos cada cuarto de hora. Se ha decidido que una buena forma de aprovechar la entrada de datos es el análisis de la variable CO2, la cual se mide en PPM.

Según el previo estudio de los registros iniciales, se concluyó que un registro de más de 600PPM en cuanto a CO2 se considera ya excesivo y, por lo tanto, perjudicial para la salud. De esta conclusión se extrae la idea de la creación de un script de Python en el cual, cuando los niveles de CO2 de esa aula superan la cantidad mencionada, automáticamente se envía un correo al email raíz del aula para que el profesor que

esté en ese momento pueda advertir el peligro y tome la decisión de abrir las ventanas con el fin de reducir el dióxido de carbono en el ambiente.

Cabe recordar que las consultas de PostgreSQL a los datos se realizan mediante tareas a intervalos planificados en Airflow cada 15 minutos, es por ello por lo que los avisos se pueden producir en este intervalo.

De esta manera se asegura que las aulas no estén constantemente recibiendo correos sino, como mucho, uno cada cuarto de hora con motivo preventivo frente al exceso de CO2.

H2. Observar los niveles térmicos y partículas en suspensión de cada aula

Durante la época de la pandemia del COVID-19 se ha puesto en valor todavía más la importancia de controlar algunas medidas de calidad del aire como son las partículas en suspensión. En función de su tamaño, éstas pueden ser gravemente perjudiciales para la salud, llegando a ocasionar efectos severos e incluso irreversibles en el sistema respiratorio.

Este hito se trata de una decisión que se está tomando más a largo plazo, pero lo que se pretende es igualar la calidad ambiental en todas las aulas. Por ello, gracias al análisis llevado a cabo en Excel sobre partículas en suspensión, se definirá qué aulas necesitan filtros de aire para mantener la calidad mencionada.



Figura 3 Cuadro de mando Power BI I

Se pretende esperar a obtener un histórico de un año de datos para contar con una realidad objetiva de la ventilación de cada aula y, en función de ello, tomar la decisión de instalar nuevos sistemas de filtrado y renovación del aire.

H3. Definir mediante diferentes medidas de agregación la situación ambiental de las aulas

Consiste en establecer un balance de situación en el cual se pueda tener una perspectiva de los niveles de cada parámetro en cada aula a lo largo de una línea temporal.

Mediante Power BI y Excel, se ha conseguido que estas visualizaciones sean accesibles y puedan ser compartidas con una gran facilidad.

No solamente consiste en ver datos numéricos, sino gráficos y tablas donde se puede consultar el dato deseado con una respuesta rápida y precisa.

Para permitir que el colegio cuente de inmediato con estos datos y visualizaciones, se han habilitado dos pantallas en diferentes áreas del colegio que pueden ser consultadas con una actualización horaria.

Los cuadros de mando a los que pueden acceder desde estas pantallas tienen el aspecto que se muestra en la Figura 3.

Como se puede observar en dicha figura, el cuadro se compone de un mapa en el que aparecen las aulas que componen cada planta con sus indicadores. En este caso, una media diaria de CO2, humedad, partículas en suspensión y temperatura. Además, se han definido gráficas a nivel horario para poder ver la

evolución de las variables, así como medias globales por planta.

De cara al futuro se prevé que estas visualizaciones sean accesibles a través de su propia web con este mismo horario de actualización programado.

H4. Buscar la relación existente entre CO2 y Temperatura

En una primera vista teórica se ha planteado la posibilidad de que el CO2 esté correlacionado con la temperatura. La hipótesis se basa en que, a mayor temperatura exterior, antes se abren las ventanas, por lo que antes desaparece el CO2 y, al contrario, cuando más frío hace, más se tardaría en abrir las ventanas y, por lo tanto, más CO2 en el aula.

Como se puede observar, se está tratando la temperatura exterior, no la interior. Por ello, se ha acudido a las bases de datos de AEMET, para poder buscar correlación entre temperatura ambiente fuera de las aulas y CO2 dentro de las mismas.

El resultado de este análisis no ha sido el esperado, ya que el resultado de la correlación ha sido prácticamente nulo, descartando la hipótesis inicial que se había planteado.

Aun así, se ha permitido comparar las temperaturas externas a las aulas con las internas. Estas visualizaciones se han dejado también a disposición del público en el Power BI ya que se han considerado de interés.

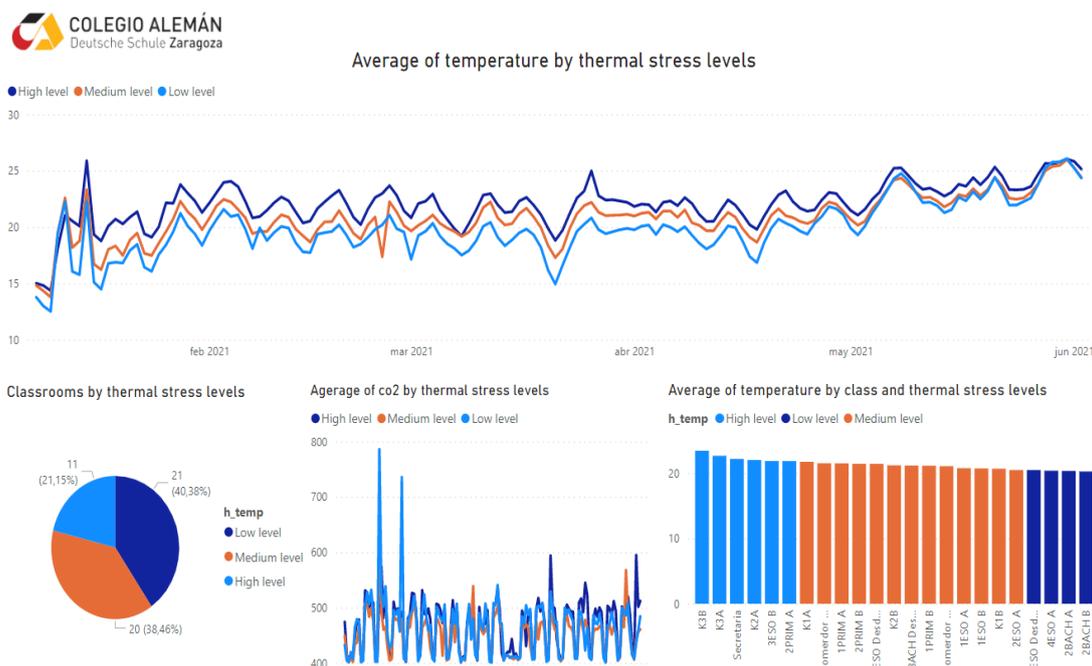


Figura 4 Cuadro de mando Power BI II

H5. Encontrar diferencias de saturación térmica en las aulas

El concepto de saturación térmica hace mención a las aulas cuyos niveles de temperatura son más elevados de lo normal.

Gracias a la colaboración de los profesores, se ha advertido que podría haber aulas con importantes diferencias de temperatura. Antes de tomar decisiones al respecto, se ha querido demostrar objetivamente que este caso es verídico y real.

Tras un estudio de las ubicaciones de las aulas, se ha definido que existen tres grupos de aulas: las que cuentan con temperaturas normales (orientación oeste), las que tienen un estrés térmico medio (orientación este) y las que tienen un estrés térmico alto (orientación sur).

El análisis se ha llevado a Power BI donde se observó que, claramente, estos grupos de aulas cuentan con temperaturas totalmente distintas y que realmente existe el problema planteado.

El dashboard resultante del estudio se muestra en la Figura 4.

En la Figura 4 se ven claramente los tres grupos de aulas que se han mencionado anteriormente y sus temperaturas a lo largo del tiempo. Se puede observar cómo en función del grupo térmico al que pertenece cada clase, hay una variación bastante elevada en cuanto a temperaturas.

Gracias a este estudio se han podido tomar medidas en cuanto a refrigeración de aulas a distintos niveles, puesto que no todas necesitan el mismo nivel de acondicionamiento térmico.

Con estos dos hitos, se ha llevado a cabo el último de los objetivos que pretende cumplirse con el proyecto.

4.3. Dashboards

Como ya se ha mencionado anteriormente, el proyecto tiene definidos numerosos hitos relativos a la infraestructura y arquitectura de datos, pero por otro lado, su exigencia en cuanto a análisis, extracción de resultados y conclusiones, ha conllevado la generación de interesantes informes de diferentes características.



Figura 5 Cuadro de mando Grafana

Estos informes son los que finalmente han ayudado a la institución a comprender su situación ambiental, sus carencias o necesidades y a ponerles solución.

Por ello, se ha visto relevante dedicar un apartado a las diferentes herramientas utilizadas, así como dashboards resultantes con cada una de ellas:

4.3.1. Herramienta Grafana

El primer informe que se ha llevado a cabo desde el inicio del proyecto fue con el software libre de Grafana.

Se trata de una herramienta enfocada a la visualización de datos de series temporales, lo cual para el análisis a largo plazo ha resultado realmente útil.

Desde un primer momento fue conectado a la base de datos y generaba por sí solo una gran cantidad de contenido realmente interesante, de carácter analítico y visual.

Como se puede observar en la Figura 5, en este cuadro aparecen los datos actuales de temperatura y CO2 por aula. El diseño de los gráficos haciendo uso de diferentes colores en función de las medidas tomadas, permite advertir en un simple golpe de vista las aulas que más (o menos) CO2 o temperatura tienen en ese momento.

Para entrar en profundidad en cuanto al tratamiento previo de datos y análisis posterior de los mismos, esta herramienta se queda limitada, por lo que se decidió recurrir a Excel y Power BI para este cometido.

4.3.2. Herramienta Power BI

Una parte realmente importante del proyecto en lo relativo al análisis ha sido la ETL, es decir, el previo tratamiento de los datos con la finalidad de llegar al análisis deseado.

Power BI cuenta con las ventajas de PowerQuery que permite realizar tratamiento de datos en M de manera rápida y eficaz.

En numerosas ocasiones se ha hecho uso de este lenguaje para preprocesado durante el proyecto. Por ejemplo, el análisis de la saturación térmica que ha requerido de etiquetas y agrupación de datos o la generación de un calendario propio para poder elaborar filtros temporales.

Además del análisis previo de los datos, tanto Power BI como Excel, son especialmente buenos para tratar modelos relacionales como el diseñado en la base de datos que contaba con tablas de hechos y dimensiones.

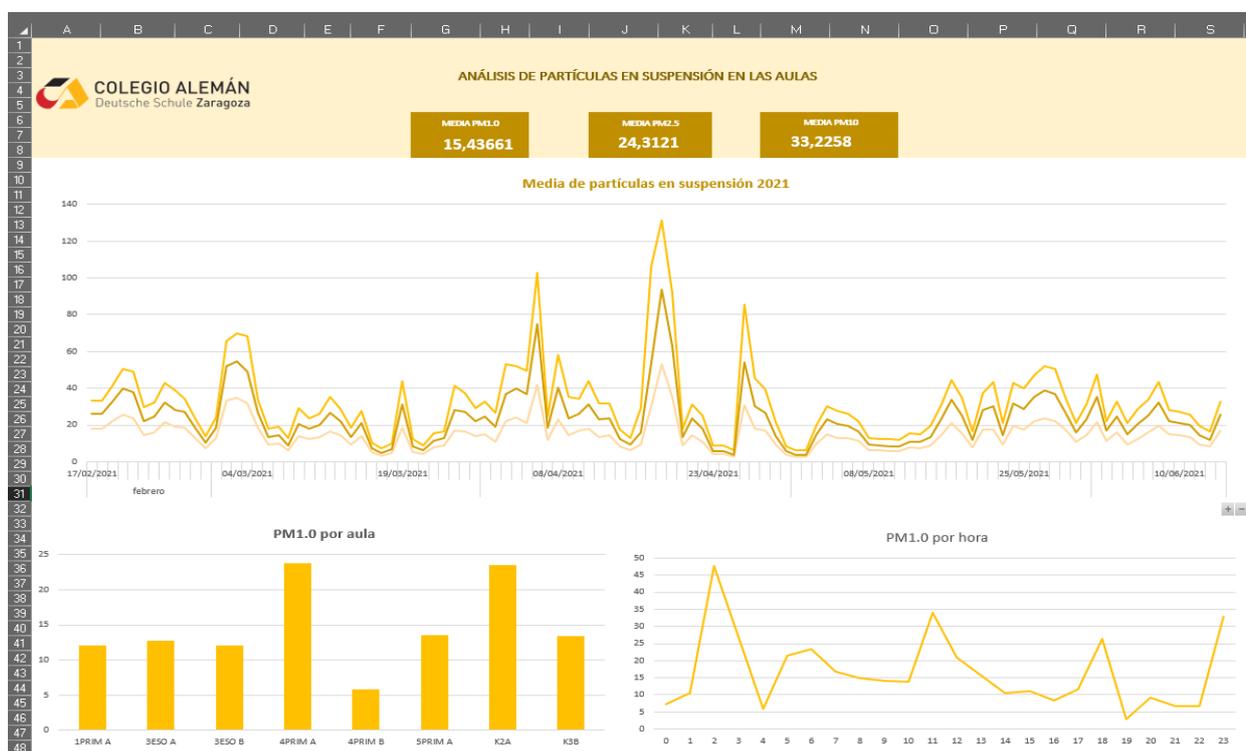


Figura 6 Cuadro de mando Excel

Gracias a la interfaz de estos programas, se pueden rápidamente relacionar todas las tablas y proporcionar visualizaciones a partir de las mismas sin necesidad de aumentar las dimensiones de las tablas, lo cual provocaría un coste computacional alto. Podemos ver en la figura 7 un claro ejemplo de modelo relacional básico que se ha utilizado para visualizar datos.

Por otra parte, la actualización de la base de datos en la nube ha sido otra de las funcionalidades que se han aprovechado de esta herramienta. La capacidad de actualizar ocho veces al día ha permitido que el colegio cuente con datos actuales accediendo desde su propia plataforma.

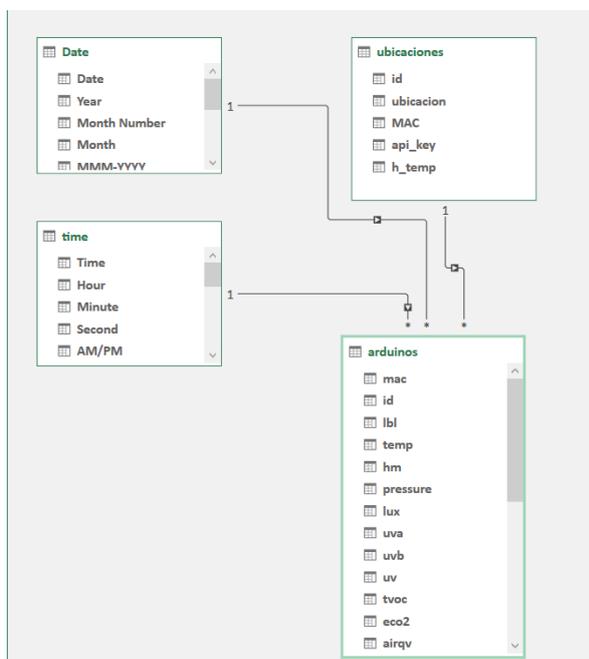


Figura 7 Modelo de datos relacional

4.3.3. Herramienta Excel

Para el caso de estudio, Excel ha realizado una labor bastante similar a Power BI, pero sin la función de actualización en la nube. En este caso, se ha optado por un fichero a nivel de escritorio que será manejado por los profesores y administración del colegio a nivel interno puesto que define el análisis de partículas en suspensión.

Como se puede ver en la Figura 6, la herramienta permite, a parte de la modelización y arquitectura de datos, el diseño de dashboards que pueden ser expor-

tados en diferentes formatos según se necesite. Excel es una aplicación muy conocida en el mundo educativo y también ha sido ese uno de los motivos por los cuales se ha elegido como representador de datos.

En este caso en concreto, se ha llevado a cabo el análisis de las partículas en suspensión en el aire por aula, hora y día. Este cuadro de mando ha permitido evaluar qué tan efectivas han sido las medidas que ha ido tomando el colegio para reducir este indicador. De hecho, se puede apreciar que, en los últimos meses de análisis, ha bajado en gran medida.

5. Conclusión y futuro trabajo

En cualquier institución o entidad que genere un mínimo de datos, es interesante analizarlos con la finalidad de entender bien el entorno y poder actuar en consecuencia.

En el caso de este trabajo, la recopilación y posterior análisis de los datos ha servido para tomar decisiones acertadas mejorando la calidad de vida de las personas que trabajan o estudian en las instalaciones.

Se trata de un colegio que, tras haber puesto en marcha esta iniciativa, no ha confinado ninguna de sus aulas debido al COVID-19. Asimismo, ninguno de sus alumnos ni profesores se ha contagiado del virus en el colegio. Cabe destacar que la toma de decisiones debe ir de la mano de acciones reales. Con esto se quiere destacar que enseguida que se ven unos niveles de CO2 altos o se advierte que existen algunas aulas que sufren un estrés térmico, se han tomado medidas rápidas y eficaces gracias a los datos.

En lo relativo a la infraestructura, se puede decir que una estructura completa no tiene por qué ser compleja, es decir, puede ser algo a lo que todo el mundo pueda acceder y, sobre todo, pueda aprender y llevar a cabo. Si bien es cierto que existen muchos aspectos técnicos que valorar, lo que se ha pretendido es crear un *Data Warehouse* para “todos”. De hecho, uno de los objetivos futuros que se plantea el propio colegio es que empiecen a ser los alumnos quienes mantengan la base de datos, puedan escribir consultas básicas e incluso exploten los datos con sus propios objetivos de investigación.

Por otra parte, en cuanto a posibles trabajos futuros a raíz de este proyecto, cabe mencionar que el trabajo ya estaba diseñado para conformar unas bases sobre las cuales poder realizar posteriores análisis útiles para la institución. Es por ello por lo que ahora que se cuenta con una infraestructura sólida, se pretende que ésta sea utilizada para otros fines de interés. El análisis de calidad del aire es solo un ejemplo de la cantidad de mejoras que se pueden realizar. Un ejemplo de análisis que se consideraría interesante sería el estudio de los niveles académicos en función de las notas y evolución de los alumnos. Su desempeño podría ser medido y se podría llevar un seguimiento personalizado de cada alumno (y de todos en global), viendo su evolución en su etapa escolarizado en el centro.

Como este análisis, se podrían llevar a cabo muchos otros alimentando bases de datos con fuentes nuevas o existentes ganando autonomía y aprovechando toda la información generada para la toma de decisiones.

Por otro lado, de cara al futuro análisis con la base de datos de calidad del aire, se pretende buscar variables que estén correlacionadas con los datos para poder sacar más información y poder introducir algoritmos de Machine Learning que sí que extraigan resultados coherentes y de interés.

Referencias

- [1] T.L.M. (8 de agosto de 2021). Los colegios se enfrentan al examen de la digitalización. EL PAIS. Obtenido de <https://elpais.com/mamas-papas/2021-03-08/los-colegios-se-enfrentan-al-examen-de-la-digitalizacion.html>
- [2] Sebastián, P. (5 de mayo de 2020). Sanidad pide a los colegios que estén preparados por si los niños vuelven a las aulas. República.
- [3] Alonso, C. (26 de febrero de 2021). La digitalización en las aulas, una tarea que sigue pendiente. Diari de Tarragona.
- [4] Cloudino. (s.f.). Obtenido de <http://semanticwebbuilder.org.mx/swb/DocumentacionSWB/PlataformaArduino>
- [5] Crespo, E. (5 de marzo de 2020). Aprendiendo Arduino. Obtenido de <https://aprendiendoarduino.wordpress.com/2020/03/05/que-es-node-red/>
- [6] Solé, R. (2021). Raspberry Pi: Crea proyectos DIY por muy poco dinero. Profesional review.
- [7] Cuezva, B. (diciembre de 2020). Introducción a Bases de Datos. Zaragoza.
- [8] Anónimo. (2020). *Red Hat*. Obtenido de <https://www.redhat.com/es/topics/automation/what-is-orchestration>