



Universidad de
Castilla-La Mancha

Producto Big Data

SmartPool

Autor: Iván Vicente Hernández García de Mora

Estudios: Máster Universitario en Ingeniería Informática

Curso: 2025/26

Asignatura: Procesamiento Masivo de Datos

Fecha: 16 de diciembre del 2025

Índice

1	Introducción y contexto del producto	3
2	Caso de uso y usuarios objetivo	3
3	Retos y preguntas clave de análisis	4
3.1	Calidad del agua y prevención de incidencias	4
3.2	Uso de la depuradora y consumo energético	5
3.3	Influencia del clima y del uso en el comportamiento de la piscina	5
3.4	Planificación del mantenimiento y priorización de intervenciones	5
4	Fuentes de datos y requisitos Big Data	6
4.1	APIs propias de sensorización de piscina	6
4.2	Datos meteorológicos de AEMET OpenData	7
4.3	Datos meteorológicos globales: Open-Meteo en AWS	7
4.4	Precios de la electricidad	8
4.5	Registros de mantenimiento y uso de la piscina	8
5	Dimensiones Big Data: las cinco V	9
5.1	Volumen	9
5.2	Variedad	9
5.3	Velocidad	10
5.4	Veracidad	10
5.5	Valor	10
6	Procesamiento de datos: <i>streaming</i> y <i>batch</i>	11
7	Arquitectura lógica de la solución	11
8	Valor esperado y posibles extensiones	12

1. Introducción y contexto del producto

En esta propuesta de proyecto se presenta *SmartPool*, un producto Big Data orientado a la supervisión y mantenimiento inteligente de piscinas privadas, comunitarias y de alojamientos turísticos. La idea principal es aprovechar datos procedentes de sensores IoT y de distintas fuentes externas, como servicios meteorológicos y precios de la energía, para conocer mejor el estado de la piscina, anticipar problemas y reducir el esfuerzo asociado a su cuidado.

En muchos casos, el mantenimiento de una piscina implica desplazarse con frecuencia hasta la instalación para medir manualmente parámetros como el pH, el cloro, la temperatura del agua o el nivel de llenado, además de gestionar la depuradora y las válvulas de entrada de agua. Esto consume tiempo, requiere cierta forma física y puede resultar especialmente complicado cuando la persona responsable tiene limitaciones de salud o la piscina se encuentra en una segunda residencia alejada del domicilio habitual. En ese contexto, disponer de un sistema que permita vigilar y tomar decisiones a distancia puede marcar una diferencia importante en términos de comodidad y calidad de vida.

SmartPool se sitúa en el punto medio entre domótica, Internet de las Cosas (IoT) y análisis de datos a gran escala. A partir de un conjunto de sensores instalados en la piscina (por ejemplo, para medir pH, cloro, temperatura y nivel de agua), el sistema genera un flujo continuo de información que se enriquece con datos externos de predicción meteorológica y precios horarios de la electricidad. Con esta combinación es posible modelar el comportamiento de la piscina como un sistema dinámico y estudiar cómo responden sus parámetros a factores como el clima, el uso o las decisiones de mantenimiento.

Desde el punto de vista del procesamiento de datos, el proyecto se plantea como un *producto Big Data* que integra datos de distinta naturaleza y velocidad:

Por un lado, se trabajará con datos en *streaming* procedentes de los sensores de la piscina, que se procesarán de forma continua para generar alertas, indicadores de estado y recomendaciones a corto plazo. Por el otro, se utilizarán datos en modo *batch*, incluyendo históricos de clima, precios de la electricidad y registros de mantenimiento, con el objetivo de identificar patrones estacionales, estimar consumos y explorar estrategias de optimización.

2. Caso de uso y usuarios objetivo

El caso de uso de *SmartPool* nace de una experiencia personal: una piscina que no se encuentra en la vivienda principal, sino en un pueblo o segunda residencia a varios kilómetros de distancia, y cuya supervisión requiere desplazamientos periódicos. La persona que se ocupa del mantenimiento debe comprobar presencialmente el estado del agua, ajustar el pH y el cloro, revisar el nivel de agua y decidir cuándo activar la depuradora.

Cuando esos desplazamientos se complican por cuestiones de tiempo, distancia o salud (como en este caso), las tareas de mantenimiento conllevan un esfuerzo mayor y tienden a espaciarse, lo que acaba provocando el deterioro de la piscina. A partir de esta necesidad, el sistema se plantea como una plataforma que centraliza la información de la instalación y permite tomar decisiones basadas en datos, así como ejecutar determinadas acciones mediante actuadores (aunque esta última parte está más cerca del ámbito de los sistemas ciberfísicos y podría quedar fuera del foco principal del trabajo).

La idea es combinar medidas continuas de sensores instalados en la piscina con información externa de clima y precios de la electricidad, de modo que el sistema pueda responder tanto a preguntas inmediatas (“¿cómo está la piscina ahora mismo?”, “¿cuál es el consumo actual?”, “¿tiene sentido encender la depuradora si va a llover?”) como a cuestiones de planificación (“¿en qué franja horaria debería activar la depuradora para gastar menos?”).

Los principales usuarios objetivo del producto son:

- **Propietarios de piscinas privadas** que quieren supervisar y mantener su instalación con mayor comodidad y precisión, ya se trate de una piscina en la vivienda habitual o en una segunda residencia, reduciendo visitas innecesarias y el riesgo de encontrarse el agua en mal estado.
- **Empresas de mantenimiento de piscinas** que gestionan varias instalaciones repartidas en distintas ubicaciones y necesitan priorizar intervenciones, planificar rutas y anticipar problemas antes de que aparezcan quejas o incidencias graves.
- **Propietarios de alojamientos turísticos** (casas rurales, hoteles con piscina, apartamentos turísticos o viviendas anunciadas en plataformas como Airbnb) interesados en mantener unas condiciones óptimas de la piscina durante toda la temporada, con una visión clara del impacto en consumo de productos químicos y en la factura eléctrica.
- **Instalaciones deportivas y polideportivos con piscinas climatizadas**, donde el consumo energético asociado a la calefacción del agua y a la filtración es especialmente relevante y resulta útil contar con herramientas que ayuden a ajustar horarios y temperaturas de forma más eficiente.

A medio y largo plazo, el sistema almacena un histórico de datos de varias temporadas que permite analizar el comportamiento de cada piscina en función del clima, del uso y de las decisiones de mantenimiento. Esto abre la puerta a responder preguntas como qué estrategias de filtrado son más eficientes, cómo se distribuye el consumo energético a lo largo del año o en qué momentos se concentran los problemas de calidad del agua. El objetivo no es solo facilitar la gestión del día a día, sino también ofrecer una base objetiva para mejorar la planificación del mantenimiento y reducir el impacto económico y operativo asociado a la piscina.

El uso típico del sistema sería el siguiente: los sensores de la piscina envían continuamente lecturas de parámetros como pH, cloro, temperatura y nivel de agua. Estos datos se reciben en la plataforma en forma de flujo (*streaming*) y se combinan con información meteorológica actual y prevista, así como con precios horarios de la electricidad. A partir de esta información, *SmartPool* calcula indicadores de estado (por ejemplo, riesgo de que el cloro caiga por debajo del umbral recomendado), genera alertas cuando se detectan valores críticos y propone recomendaciones de actuación, como activar la depuradora en determinadas franjas horarias o programar el vertido de productos químicos.

3. Retos y preguntas clave de análisis

SmartPool se diseña para responder a un conjunto de preguntas que combinan la supervisión inmediata de la piscina con el análisis de su comportamiento a lo largo del tiempo. A continuación se recogen los principales retos que guiarán el diseño del producto y el uso de las distintas fuentes de datos.

3.1. Calidad del agua y prevención de incidencias

Uno de los objetivos centrales del sistema es mantener la piscina dentro de unos rangos de calidad aceptables sin necesidad de supervisión constante en persona. En este aspecto, las preguntas clave son:

- ¿En qué momento la piscina corre mayor riesgo de salirse de los rangos recomendados de pH y cloro, teniendo en cuenta el estado actual del agua y la previsión meteorológica de las próximas horas o días?
- ¿Se pueden detectar de forma temprana patrones que suelen preceder a problemas típicos como agua verde, exceso de turbidez o algas?
- ¿Cuánto tiempo ha permanecido la piscina fuera de los umbrales definidos de seguridad en la última semana o en la última temporada?

Responder a estas preguntas requiere combinar las lecturas actuales de los sensores con datos históricos de la propia piscina y de las condiciones meteorológicas en temporadas previas.

3.2. Uso de la depuradora y consumo energético

El segundo bloque está relacionado con la optimización de la depuradora y el impacto en la factura eléctrica. En este caso, el sistema se plantea preguntas como:

- ¿Cuáles son las mejores horas del día para activar la depuradora, equilibrando el coste de la energía con la necesidad de mantener el agua en condiciones óptimas para el baño?
- ¿Cómo se reparte el consumo energético de la depuradora a lo largo de un día, de una semana o de toda la temporada de uso de la piscina?
- ¿Qué efecto tendría adelantar o retrasar determinadas horas de filtrado en función de los precios horarios de la electricidad y de la previsión de uso de la piscina?
- En piscinas climatizadas, ¿cómo influye el uso del sistema de calefacción (por ejemplo, bombas de calor o calderas) en el consumo energético total y cómo se puede ajustar la temperatura objetivo del agua en función de la previsión meteorológica y del precio esperado de la electricidad?

Para responder a estas cuestiones utilizaremos tanto datos de sensores (estado de la bomba de filtrado, tiempo de funcionamiento, caudal aproximado, temperatura del agua y, en su caso, estado del sistema de calefacción) como series históricas y casi en tiempo real de precios de la electricidad, lo que permite analizar diferentes escenarios de programación y su impacto económico.

3.3. Influencia del clima y del uso en el comportamiento de la piscina

La calidad del agua no depende únicamente de los tratamientos aplicados, sino también del contexto climático y del uso que se hace de la piscina. Por ello, *SmartPool* plantea retos como:

- ¿Cómo influyen las olas de calor, los episodios de lluvia o los cambios bruscos de temperatura ambiente en la evolución del pH, del cloro y de la temperatura del agua?
- ¿Qué diferencias se observan entre distintas épocas del año, por ejemplo entre el inicio de la temporada de baño, los meses de máximo uso y los periodos de menor actividad?
- ¿Cómo se relaciona la intensidad de uso de la piscina (número de bañistas, frecuencia de uso) con el aumento de consumo de productos químicos y con la aparición de incidencias de calidad?

Para poder abordar estas preguntas, la plataforma combinará datos meteorológicos procedentes de fuentes abiertas con registros de uso y con el histórico de sensores de la piscina.

3.4. Planificación del mantenimiento y priorización de intervenciones

Finalmente, el sistema también pretende servir como ayuda a la planificación del mantenimiento, especialmente en escenarios en los que una misma persona o empresa gestiona varias piscinas en paralelo. En este ámbito, las preguntas clave serían:

- ¿Qué piscinas presentan un mayor riesgo actual o potencial de incidencia y deberían recibir prioridad en la próxima intervención presencial?
- ¿Qué estrategias de mantenimiento (horas de filtrado, frecuencia de adición de productos, ajustes manuales) han demostrado ser más eficaces a lo largo de varias temporadas, tanto en términos de calidad del agua como de coste?
- ¿Es posible proponer, a partir de los datos, una planificación de mantenimiento que reduzca desplazamientos innecesarios y concentre las intervenciones en los momentos en los que aportan mayor beneficio?

4. Fuentes de datos y requisitos Big Data

El diseño de *SmartPool* se fundamenta en varias fuentes de datos heterogéneas que aportan información complementaria sobre el estado de la piscina, las condiciones externas y las decisiones de mantenimiento. Estas fuentes se combinan en dos modos de trabajo principales: datos en *streaming*, orientados a la monitorización casi en tiempo real, y datos en *batch*, pensados para el análisis histórico y la planificación a medio y largo plazo.

4.1. APIs propias de sensorización de piscina

Como base del sistema se plantea una capa de servicios en Python para simular y exponer los datos procedentes de la piscina. Estos servicios podrían implementarse, por ejemplo, con **FastAPI** u otro framework similar, actuando como pasarela entre los sensores físicos (o su simulación) y la plataforma de procesamiento.

El objetivo de estas APIs es generar y publicar eventos que representen lecturas periódicas de sensores, así como proporcionar accesos de tipo *batch* a históricos de medidas. En el contexto de la asignatura, este enfoque se alinea con las APIs de ejemplo que se proporcionan para simular datos de temperatura y humedad; en caso de que las APIs de los profesores no fueran suficientes o no se adaptaran bien al escenario concreto de la piscina, se ampliaría su funcionalidad o se desplegaría un servicio propio con la misma filosofía.

De forma orientativa, las lecturas de cada piscina incluirían atributos como:

- **timestamp**: marca temporal de la lectura.
- **id_piscina**: identificador de la piscina o instalación.
- **temperatura_agua**: temperatura del agua en grados.
- **ph**: nivel de pH estimado.
- **cloro_mg_l**: concentración de cloro.
- **nivel_agua**: nivel relativo de agua (por ejemplo, en centímetros o porcentaje).
- **estado_bomba**: indicador de si la depuradora está activada o no.

Velocidad de los datos:

- *Streaming*: eventos de sensor enviados con una frecuencia configurable (por ejemplo, cada 30–60 segundos), que alimentan la parte de monitorización y alertas en tiempo casi real.
- *Batch*: volcados periódicos (por ejemplo, diarios) de las lecturas acumuladas, almacenados para análisis históricos y construcción de modelos de comportamiento de la piscina.

4.2. Datos meteorológicos de AEMET OpenData

El comportamiento de la piscina está fuertemente condicionado por el clima local. Para capturar este efecto, se plantea utilizar datos meteorológicos procedentes de *AEMET OpenData*, que ofrece información histórica y casi en tiempo real para diferentes estaciones a lo largo del territorio.

A partir de esta fuente se podrán consumir, por ejemplo:

- Registros históricos de temperatura del aire, humedad relativa, precipitación y radiación.
- Datos recientes y predicciones a corto plazo para la zona donde se ubica la piscina.

Velocidad de los datos:

- *Batch*: descarga de históricos para analizar temporadas completas y estudiar el impacto del clima en la calidad del agua y el consumo de productos.
- *Micro-batch*: consultas periódicas (por ejemplo, cada hora) a la API para obtener clima actual y pronósticos, que pueden integrarse con el flujo de sensores para refinar las alertas y recomendaciones.

4.3. Datos meteorológicos globales: Open-Meteo en AWS

Además de la información meteorológica local, el proyecto puede utilizar conjuntos de datos globales para reforzar su carácter Big Data y facilitar futuros escenarios de expansión. En este sentido, se plantea el uso del *Open-Meteo Weather API Database*, disponible como conjunto de datos abierto dentro de la Amazon Sustainability Data Initiative.

Este recurso proporciona series temporales históricas y de predicción basadas en distintos modelos meteorológicos y accesibles a través de la infraestructura de AWS. En el contexto de *SmartPool*, su papel sería principalmente:

- Aportar datos meteorológicos históricos de largo plazo para distintas ubicaciones, permitiendo experimentar con escenarios de despliegue del producto en piscinas situadas en otras regiones o países.
- Servir como base para análisis comparativos entre diferentes climas y para la construcción de modelos que no se limiten únicamente a un conjunto reducido de estaciones locales.

Aunque en la fase inicial el foco esté en una piscina concreta, contar con esta fuente facilita escalar el producto a un contexto de muchas instalaciones repartidas geográficamente.

4.4. Precios de la electricidad

Para poder optimizar el uso de la depuradora desde el punto de vista económico, resulta útil disponer de series históricas y actuales de precios de la electricidad. En el caso de España, una opción natural es trabajar con datos de la tarifa regulada (por ejemplo, PVPC) publicados a través de plataformas oficiales y accesibles mediante API, así como con conjuntos históricos ya procesados.

Los datos relevantes incluirían, entre otros:

- **fecha y hora:** identificación del intervalo horario.
- **precio_kwh:** precio de la energía correspondiente a cada hora.
- Información sobre el tipo de periodo (punta, llano, valle), si procede.

Velocidad de los datos:

- *Batch:* históricos de precios para analizar el coste acumulado de la depuradora, estudiar variaciones estacionales y simular estrategias de programación alternativas.
- *Micro-batch:* consultas recurrentes a la API para conocer los precios horarios del día siguiente y ajustar la planificación de filtrado.

4.5. Registros de mantenimiento y uso de la piscina

La parte de negocio del sistema se alimentará con datos procedentes del propio mantenimiento de la piscina y del uso real que se hace de ella. Estos datos podrán almacenarse en ficheros o bases de datos y se gestionarán en modo *batch*.

Algunos campos ilustrativos serían:

- **Mantenimiento:**
 - **fecha e id_piscina.**
 - **tipo_intervencion** (adición de cloro, corrector de pH, lavado de filtro, ajuste de nivel, etc.).
 - **cantidad_producto** y observaciones relevantes.
- **Uso de la piscina:**
 - **fecha e id_piscina.**
 - Indicadores de **intensidad_uso** (número estimado de bañistas, horas de uso).
 - Eventos asociados (por ejemplo, reservas en alojamientos turísticos).

Estos datos permiten relacionar decisiones de mantenimiento y patrones de uso con el estado del agua, el consumo energético y la aparición de incidencias.

5. Dimensiones Big Data: las cinco V

El diseño de *SmartPool* se ha planteado teniendo en cuenta las cinco V clásicas asociadas al Big Data: volumen, variedad, velocidad, veracidad y valor. En esta sección se resume cómo el producto aborda cada una de estas dimensiones.

5.1. Volumen

Aunque el caso de uso puede partir de una única piscina, la propuesta está pensada para escalar a escenarios con muchas instalaciones gestionadas de forma centralizada. El volumen de datos se genera principalmente por:

- Lecturas de sensores de cada piscina con una frecuencia elevada (por ejemplo, cada pocos segundos o minutos), lo que produce grandes series temporales a lo largo de toda la temporada de baño.
- Históricos meteorológicos de AEMET y de Open-Meteo en AWS, que pueden abarcar varios años o décadas de datos para diferentes ubicaciones.
- Series históricas de precios horarios de la electricidad, con registros diarios continuos.
- Registros de mantenimiento y uso acumulados a lo largo de varias temporadas.

Al combinar todas estas fuentes y considerar un número creciente de piscinas, el sistema debe estar preparado para manejar volúmenes de datos significativos, tanto en almacenamiento como en procesamiento.

5.2. Variedad

SmartPool integra datos de naturaleza diversa, que capturan distintas facetas del comportamiento de la piscina y de su entorno:

- Datos de sensores IoT expuestos a través de APIs propias en Python, que describen el estado físico de la piscina (pH, cloro, temperatura del agua, nivel, estado de la depuradora).
- Datos meteorológicos locales procedentes de AEMET OpenData, con información sobre temperatura del aire, humedad, precipitación, radiación o viento.
- Datos meteorológicos globales de Open-Meteo en AWS, útiles para análisis comparativos y para plantear escenarios de despliegue en otras regiones.
- Series temporales de precios de la electricidad, que añaden una dimensión económica al problema.
- Registros de mantenimiento y uso de la piscina, que recogen decisiones humanas, intervenciones manuales y patrones de ocupación.

Esta combinación mezcla datos físicos, climatológicos, económicos y de negocio, lo que obliga a definir modelos de datos flexibles y procesos de integración adecuados.

5.3. Velocidad

La solución distingue entre diferentes ritmos de actualización:

- **Alta velocidad:** flujo continuo de lecturas de sensores que llegan en forma de eventos en streaming y permiten monitorizar el estado de la piscina casi en tiempo real, generando alertas cuando se superan ciertos umbrales.
- **Velocidad intermedia:** consultas periódicas a APIs externas para obtener clima reciente o precios de la electricidad del día siguiente, que se integran en forma de micro-batch con el flujo de sensores.
- **Velocidad baja:** procesos batch diarios o semanales que recorren los históricos para calcular métricas agregadas, entrenar modelos y generar informes.

De este modo, el sistema combina procesamiento en streaming para la supervisión inmediata con procesamiento batch para análisis más pesados y de largo plazo.

5.4. Veracidad

La veracidad hace referencia a la calidad y fiabilidad de los datos. En este proyecto aparecen varios retos:

- Los sensores de la piscina pueden producir lecturas ruidosas o erróneas debido a problemas de calibración, suciedad, cortes de alimentación o errores de comunicación.
- Los datos meteorológicos proceden de diferentes fuentes, con resoluciones espaciales y temporales distintas, por lo que es necesario armonizarlos y detectar posibles incoherencias entre AEMET y los conjuntos globales de Open-Meteo.
- Las series de precios de la electricidad y los registros de mantenimiento pueden contener valores faltantes o errores de registro que conviene depurar.

Para abordar estos problemas, la arquitectura contempla una zona de datos en bruto y una zona de datos depurados, donde se aplican técnicas de limpieza, detección de outliers y validación cruzada entre fuentes. Además, el uso de fuentes oficiales para clima y precios ayuda a mejorar la confianza en la información externa que se incorpora al sistema.

5.5. Valor

La última V está relacionada con el impacto que se obtiene a partir de los datos. En este proyecto, ese valor se centra en poder tomar decisiones más precisas y oportunas sobre el mantenimiento de la piscina, reduciendo la improvisación y basándose en evidencias.

En conjunto, las cinco V muestran que SmartPool no se limita a registrar mediciones puntuales, sino que explota grandes volúmenes de datos heterogéneos, generados a diferentes velocidades y sometidos a procesos de control de calidad, con el objetivo de convertir esa información en decisiones prácticas y medibles para las personas y entidades responsables del mantenimiento de la piscina.

6. Procesamiento de datos: *streaming* y *batch*

En *SmartPool* se distinguen dos formas principales de trabajar con los datos: el procesamiento en *streaming*, orientado a la monitorización casi en tiempo real, y el procesamiento en *batch*, pensado para el análisis histórico y la planificación a medio y largo plazo. Ambas capas se alimentan de las fuentes descritas en la sección 4 y se complementan entre sí.

En la parte de *streaming*, las lecturas generadas por los sensores de la piscina se envían de forma continua hacia una capa de ingesta que actúa como punto de entrada al sistema. Cada evento incluye información sobre el estado actual de la piscina (por ejemplo, pH, cloro, temperatura y nivel de agua), así como el estado de la depuradora. A partir de estos eventos, el sistema puede calcular indicadores instantáneos, como comprobar si la piscina está dentro de los rangos de seguridad, y generar alertas cuando se detectan situaciones críticas, por ejemplo un descenso brusco del nivel de cloro o un nivel de agua anómalo. Periódicamente, este flujo se enriquece con información externa consultada en micro *batch*, como los datos meteorológicos más recientes y los precios horarios de la electricidad para las próximas horas.

El procesamiento en *batch* se nutre de la acumulación de históricos tanto de los sensores como de las fuentes externas. A intervalos definidos, por ejemplo diarios o semanales, se ejecutan tareas que agregan los datos crudos y generan tablas o ficheros resumen, con métricas como el tiempo total de filtrado por día, el consumo energético estimado, la evolución de pH y cloro o el impacto de las condiciones meteorológicas en la calidad del agua. Estos procesos permiten estudiar patrones a lo largo de toda una temporada de baño o incluso de varios años, y comparar diferentes estrategias de mantenimiento.

Además, el procesamiento en *batch* hace posible construir modelos de comportamiento de la piscina, basados en los históricos, que después pueden integrarse de forma ligera en el flujo de *streaming*. Por ejemplo, un modelo entrenado con datos pasados podría estimar la probabilidad de que el agua salga de rango en las próximas 24 horas en función del estado actual, del clima previsto y de la programación de la depuradora. De este modo, las dos capas de procesamiento se refuerzan mutuamente: el *streaming* se beneficia de lo aprendido en el *batch*, y el *batch* se alimenta de los eventos que se generan de manera continua.

7. Arquitectura lógica de la solución

A continuación se presenta, de forma resumida, una arquitectura lógica que muestra cómo encajan las distintas piezas de *SmartPool* y cómo se gestionan los datos desde que se generan en la piscina hasta que se explotan en forma de indicadores, alertas e informes, sin entrar todavía en tecnologías concretas.

En la **capa de adquisición** se sitúan las APIs de sensorización de la piscina, implementadas en Python y expuestas mediante un marco ligero como **FastAPI**. Estas APIs reciben o simulan las lecturas de sensores y las publican en forma de eventos para su procesamiento en *streaming*. En esta misma capa se integran los conectores hacia las fuentes externas: clientes para consumir datos de AEMET OpenData, mecanismos de descarga o acceso a los conjuntos de datos de Open-Meteo en AWS y llamadas a las APIs de precios de la electricidad. Además, se contemplan procesos de carga de ficheros *batch* con información de mantenimiento y uso de la piscina.

La **capa de almacenamiento** actúa como un repositorio central de datos, organizada conceptualmente como un *data lake*. En ella se distinguen, al menos, tres zonas: una zona *raw*, donde se almacenan los datos tal y como llegan desde las fuentes; una zona *curated*, con datos ya limpios y transformados; y una zona de datos agregados y modelos, donde se conservan los resultados de los procesos de análisis *batch*. Esta organización facilita la repetición de procesos, la trazabilidad de los datos y la incorporación de nuevas fuentes en el futuro.

Sobre esta base se construye la **capa de procesamiento**. Por un lado, se ejecutan tareas en *streaming* que consumen los eventos generados por las APIs de sensorización, aplican reglas de negocio (umbrales, detección de anomalías básicas) y generan alertas o indicadores en tiempo casi real. Por otro lado, se lanzan trabajos *batch* que recorren los históricos para obtener estadísticas de uso, evaluar el impacto del clima o analizar el coste energético asociado a diferentes patrones de filtrado. En esta capa también se podrían integrar algoritmos de aprendizaje automático, entrenados de manera periódica sobre los datos históricos.

Finalmente, la **capa de explotación** ofrece el acceso a los resultados del sistema. Aquí se incluirían cuadros de mando para visualizar el estado actual de cada piscina, informes periódicos con la evolución de sus principales indicadores y herramientas sencillas para simular escenarios de mantenimiento, por ejemplo comparar el coste y la calidad del agua bajo distintas configuraciones de horarios de la depuradora. Esta capa es la que interactúa directamente con los usuarios finales descritos en la sección 2 y es la responsable de traducir los datos y modelos en información accionable.

8. Valor esperado y posibles extensiones

La intención con *SmartPool* es aportar valor en varios niveles. En primer lugar, facilita la supervisión en remoto de la piscina, reduciendo la necesidad de desplazamientos frecuentes para comprobar su estado. Esto resulta especialmente relevante cuando la piscina se encuentra en una segunda residencia o cuando la persona responsable del mantenimiento tiene dificultades para acudir con regularidad a la instalación.

En segundo lugar, el sistema contribuye a mejorar la calidad del agua y a reducir la probabilidad de incidencias, al permitir detectar de forma temprana situaciones de riesgo y anticipar acciones correctoras. La combinación de datos de sensores con información meteorológica y precios de la electricidad favorece la toma de decisiones más informadas, tanto en el día a día como en la planificación de la temporada.

En tercer lugar, el proyecto busca la optimización de recursos. Por un lado, el análisis del consumo energético y de los precios horarios de la electricidad ayuda a minimizar el coste asociado al funcionamiento de la depuradora o la calefacción. Por otro, el estudio de los patrones de mantenimiento y uso permite ajustar la cantidad de productos químicos empleados, evitando tanto el fraude, que puede derivar en problemas sanitarios, como el sobreuso, que incrementa el coste y el impacto ambiental.

A partir de esta base, el producto se puede extender en varias direcciones. Una posibilidad es ampliar el alcance a redes de piscinas gestionadas por una misma empresa de mantenimiento, integrando funciones de priorización de intervenciones y planificación de rutas. Otra línea de trabajo sería incorporar modelos de predicción avanzada que estimen la evolución de la calidad del agua bajo distintos escenarios de clima, uso y mantenimiento, o que sugieran automáticamente configuraciones óptimas de filtrado. Finalmente, la misma arquitectura podría adaptarse a otros entornos con características similares, como balnearios, parques acuáticos o pequeñas instalaciones deportivas, reutilizando gran parte de las fuentes de datos y de los procesos de análisis definidos para *SmartPool*.