

## SOLUCIÓN DATA LAKE ESCALABLE PARA UNA CIUDAD INTELIGENTE

Ruth G. Obregón, Coordinadora de Proyectos, CIC Consulting Informático

**Resumen:** La ciudad inteligente aparece como una gran oportunidad para gestionar de una forma eficiente el futuro de la ciudad. Nuevos modelos analíticos pueden jugar un papel clave en el ámbito urbano contribuyendo a poner en valor la información para la toma de decisiones de carácter estratégico y operativo, lo que se traduce en una mayor eficiencia operativa, reducción de costes y servicios de mejor calidad. Utilizando datos de distintos ámbitos que, analizados de forma conjunta, ayuden en la toma de decisiones con un impacto en la reducción del consumo energético, la optimización de rutas, la descongestión urbana, la minimización del consumo de recursos hídricos, la mejora del mantenimiento de las infraestructuras, el aumento de seguridad ciudadana y la gestión de los activos propios, entre otras. Partiendo de la información de los sistemas que tenga un ayuntamiento ya sea de la red de sensores u otras fuentes, se plantea una solución que tome en consideración los distintos orígenes posibles: gestión de aparcamientos y movilidad, información meteorológica, comunicaciones TETRA de los servicios de emergencias y seguridad, ya sean públicos (policía, ambulancias, bomberos, etc.) o privados (industrias químicas y petroleras, redes de metro, redes 4G/5G para patrones de movilidad, etc.). Se trata de una herramienta que permita realizar analítica del pasado y, simultáneamente, saber lo que ocurre en el momento actual y predecir lo que sucederá. Requiere muy pocos recursos frente a otras arquitecturas, siendo escalable con facilidad.

**Palabras clave:** Smart City, IoT, Data Science, Data Lake, Data Warehouse, Big Data

### ANTECEDENTES

Actualmente, la mitad de los habitantes del mundo, más de 4.000 millones, vive en zonas urbanas y, según se prevé, en 2050 el 65% de la población mundial, cerca de 6.000 millones de personas, vivirá en urbes. Las ciudades inteligentes son una solución a esta superpoblación. El crecimiento de los núcleos urbanos presenta nuevos retos y oportunidades, así como la posibilidad de replantear los ya existentes: satisfacer la demanda de recursos naturales, mejorar las infraestructuras, o convertir las ciudades en espacios más seguros y sostenibles para vivir.

En España se está trabajando desde hace tiempo en modernizar las ciudades: Barcelona, Valencia, Madrid o Santander llevan años trabajando en esta línea y se consideran referentes nacionales. Sin embargo, hay expectativas de que la tendencia se extienda a más ciudades. Como ejemplo significativo, en la ciudad de Santander se tienen monitorizados una buena parte de los contadores de agua, lo que ha permitido optimizar la presión del caudal de agua que recibe cada ciudadano, estudiando la curva de la demanda y aplicando la presión justa en cada momento. Esto no solo ha permitido dar un mejor servicio al ciudadano, sino que ha reducido de forma significativa el consumo energético para el bombeo del agua.

Además, no se pueden olvidar las infraestructuras críticas, que son aquellas instalaciones que resultan vitales para el funcionamiento de un país, dada su actividad. La posibilidad de que se produzcan desastres o emergencias de distinto tipo y naturaleza en infraestructuras críticas o el incremento de potenciales ataques, obliga a estar prevenidos. La normativa de España define como infraestructura estratégica aquella *cuyo funcionamiento es indispensable y no permite soluciones alternativas, por lo que su perturbación o destrucción tendría un grave impacto sobre los servicios esenciales*. En esta definición se podrían incluir instalaciones de varios sectores, por ejemplo: hospitales, centrales energéticas, aeropuertos, laboratorios de investigación, empresas de abastecimiento de agua o comunicaciones (satélites). Algunos de ellos transversales como las tecnologías de la información o la Administración Pública. La Ley 8/2011 regula las medidas de protección a aplicar en paralelo al Plan Nacional de Protección de las Infraestructuras Críticas, que engloba al sector público y al privado, y cuyo objetivo es *implantar a medio plazo una cultura de seguridad en la que tanto el sector privado como las Administraciones Públicas trabajen sobre parámetros homogéneos y claramente definidos en materia de protección de sus respectivos activos, logrando una coordinación de esfuerzos y una sinergia en sus objetivos*, según se explica desde el Centro Nacional para la Protección de Infraestructuras Críticas (CNPIC). Si bien los requerimientos específicos dependen del sector donde se quieren aplicar, se encuentran elementos comunes a tener en cuenta a la hora de implantar cualquier sistema de seguridad. Es primordial realizar un análisis exhaustivo de la actividad evaluando los posibles riesgos, siempre con una mirada integradora. Las redes TETRA (Trans European Trunked Radio) son sistemas móviles digitales de radio para la comunicación entre los profesionales de sectores críticos comentados como servicios de emergencias, ya sea policía, bomberos, ambulancias, etc. Las ventajas que ofrece dicha tecnología han hecho que administraciones y ciertas empresas privadas opten por ella dado que la

mayoría de los países tiene reservada una banda para comunicaciones críticas (380-400 MHz) que permite alcanzar más cobertura por cada antena instalada, lo mismo que ocurrirá cuando se usen los 700 MHz para el 5G.

## TRANSFORMACIÓN DIGITAL

Consiste en utilizar la tecnología disponible actualmente para hacer frente a los nuevos retos que plantea la sociedad actual. Una ciudad inteligente es una zona urbana que incorpora las tecnologías digitales TIC e IoT para que la administración de sus recursos se realice de una manera más eficiente y económica. Parte de la premisa de mejorar la calidad de vida de los habitantes que la componen, planteamiento que tiene que comenzar con la participación de la ciudadanía, quien debe formar parte de su diseño y desarrollo. Para los ciudadanos de las ciudades inteligentes, la tecnología es tan sólo un medio para llegar a un fin, siendo los objetivos reales el perfeccionamiento de los servicios y de la conectividad, el mantenimiento del control de los flujos de tráfico o la mejora de la experiencia en la ciudad. Las soluciones TIC son el núcleo que permitirá gestionar de manera eficiente los recursos, pero en el desarrollo de una ciudad inteligente es crucial que los ciudadanos estén implicados y comprometidos. De ahí vendrá su éxito. Por tanto, la coordinación entre las administraciones públicas y la ciudadanía es esencial.

Por otro lado, se deben optimizar los procesos de las ciudades mediante sistemas conectados en los que la información sea más accesible. A medida que aumenta el volumen de datos recopilados de los sensores, datos abiertos, dispositivos u otros puntos finales, el valor comercial potencial que puede obtener de estos datos continúa creciendo de una manera exponencial. Analizar los datos es la clave para obtener información importante y útil de todos los datos que llegan, y así poder aplicarlos a las necesidades empresariales. Sin embargo, se requiere de una arquitectura de datos de IoT que permita recopilar adecuadamente los conocimientos importantes. La clave para avanzar hacia la transformación digital es realizar un análisis, es decir, un mecanismo que obtenga conocimiento y significado a partir de datos. Tales mecanismos van desde lo simple a lo complejo, y podrían incluir el aprovechamiento de técnicas de procesamiento de señales, técnicas de Big Data o el uso de aprendizaje automático avanzado (Machine Learning).

El objetivo final es acelerar la innovación de la ciudad estableciendo mecanismos que permitan implementar estas nuevas ideas de forma escalable para solucionar los problemas existentes y explorar nuevas oportunidades a través de soluciones tecnológicas.

## DESCRIPCIÓN DE LA SOLUCIÓN

Para dotar a las ciudades de estas funcionalidades, entendemos que la estrategia correcta sería proporcionar una plataforma Data Lake compuesta por un conjunto de servicios que faciliten el Data Science, es decir, que den utilidad a los datos. Cuando todos los datos que se necesitan son visibles, se puede usar analítica descriptiva para tomar tantas decisiones como se quiera, solo mirando los hechos. El principal beneficio de un Data Lake es la centralización de fuentes, no hay una estructura rígida permitiendo absorber todo lo presente en estas. Una vez juntas, pueden ser combinadas y procesadas. En suma, contamos con la analítica del pasado, simultáneamente sabemos lo que ocurre ahora y además podemos predecir lo que está por venir.

Un enfoque similar más tradicional es el Data Warehouse, donde se pierde bastante tiempo analizando las fuentes de datos, conociendo los procesos de negocio y adecuando los datos. Como resultado se obtiene un modelo de datos estructurados rígido y pesado de alimentar que estaría preparado para la generación de informes, siendo necesario simplificarlo para excluir aquellos datos que no se van a usar y así mejorar la velocidad de lectura.

Existen varias diferencias entre un Data Lake y un Data Warehouse. Las principales son la estructura de los datos, los usuarios finales, los métodos de procesamiento y la finalidad general de los datos. Un Data Lake es un enorme conjunto de datos en bruto cuya finalidad no se ha definido todavía. Un Data Warehouse es un repositorio de datos filtrados y estructurados que ya han sido procesados para una finalidad concreta. La única similitud real entre ambos es que los dos almacenan datos. Los datos en bruto son maleables, pueden analizarse de forma rápida y favorecen el Machine Learning. Por el contrario, el riesgo de los datos en bruto es que es necesario aportar calidad a los datos y llevar a cabo medidas de gobernanza adecuadas, es decir, una disciplina para administrar riesgos, reducir costos y maximizar su valor. En el Data Warehouse se busca dar respuesta a preguntas que normalmente ya se conocen. En el Data Lake, además de responder esas mismas preguntas, es posible plantearse otras nuevas fácilmente, encontrar correlaciones, etc.

Los datos de sensores, al igual que la actividad de las redes sociales, el texto y las imágenes, son más fácilmente consumibles y almacenables con un Data Lake. Se guardan todos los datos independientemente de la fuente y la estructura. Se conservan en su forma bruta y solo se transforman cuando van a ser utilizados, necesitando muy pocos recursos frente a otras arquitecturas. Por ello, permite acceder a los datos antes de que se hayan transformado, limpiado y estructurado, mejorando la velocidad en comparación con el método tradicional de Data Warehouse. Un Data Lake conserva casi todos los datos y, aunque en un primer momento no se prevea su análisis, se podrá hacer con posterioridad, puesto que esa cualidad de dato necesario puede ser diferente en el futuro. Además, no se necesita un macroentorno, siendo escalable con facilidad.

Se pretende manejar el volumen de datos recopilados de los sensores, datos abiertos, datos meteorológicos, sistemas de información geográfica, comunicaciones TETRA de los servicios de emergencias y seguridad, ya sean públicos (policía, ambulancias, bomberos, etc.) o privados (industrias químicas y petroleras, redes de metro, redes 4G/5G para patrones de movilidad, etc.) Una vez que tenemos esta información heterogénea centralizada, se pueden responder preguntas, detectar correlaciones, causalidades y patrones no evidentes. Con esto, podemos hablar de ayuda a la toma de decisiones, lo cual tiene impacto directo en el ciudadano a través, por ejemplo, de la reducción del consumo energético, la optimización de rutas, la descongestión urbana, la minimización del consumo de recursos hídricos, la mejora del mantenimiento de las infraestructuras o el aumento de seguridad ciudadana.



Figura 1. El análisis de datos como punto de partida para una solución IoT.

Los análisis y transformaciones se suelen almacenar en el propio Data Lake y, por tanto, pasan a convertirse a su vez en datos de origen para nuevos análisis. Es posible, en fin, establecer una jerarquía sobre los datos atendiendo a distintos criterios como nivel de abstracción, permisos, confianza o grado de elaboración.

Como extra a los análisis, tras la carga en el Data Lake es posible realizar estudios en “tiempo real”, sin tener que esperar al almacenamiento de los datos ni tener que montar una segunda arquitectura. Para ello se parte de un subconjunto de los mismos datos originales y se reutilizan algunas de las herramientas de transformación.

## CONCLUSIONES

La plataforma planteada trata de utilizar la tecnología disponible actualmente para hacer frente a los nuevos retos que plantean las ciudades inteligentes, que incorporan las tecnologías digitales TIC e IoT con el fin de que la administración

de sus recursos se realice de una manera más eficiente y económica. El objetivo final es acelerar la innovación de la ciudad para solucionar los problemas existentes y explorar nuevas oportunidades a través de soluciones tecnológicas.

Se pretende manejar el volumen de datos recopilados de los sensores, datos abiertos, comunicaciones TETRA de los servicios de emergencias y seguridad, etc. como ayuda a la toma de decisiones, lo cual tiene impacto directo en la ciudadanía, quien debe formar parte de su diseño y desarrollo. La coordinación entre las administraciones públicas y la ciudadanía es esencial ya que se parte de la premisa de mejorar la calidad de vida de los habitantes que componen las ciudades inteligentes.

Nuestra herramienta logra la mayor disponibilidad ya que aporta:

1. Visibilidad en tiempo real del estado de todas las infraestructuras.
2. Visibilidad en tiempo real del recorrido de un terminal TETRA.
3. Actúa como control de seguridad en los sistemas ya que evita la duplicidad de terminales TETRA previniendo con ello cualquier tipo de ataque malicioso.
4. Optimiza los procesos y aumenta la eficiencia operacional.
5. Acelera y mejora la toma de decisiones, generando el conocimiento de negocio necesario.
6. Fortalece las capacidades analíticas y de planificación.

## AGRADECIMIENTOS

Principalmente a mi compañero Juan Antonio Losada por sus aportes y críticas constructivas en el desarrollo de esta y otras soluciones.

## REFERENCIAS

- <https://boe.es/eli/es/l/2011/04/28/8/con> (12 marzo 2019)
- <https://www.boe.es/buscar/pdf/2011/BOE-A-2011-7630-consolidado.pdf> (12 marzo 2019)
- [https://www.motorolasolutions.com/es\\_xl/productos/tetra.html](https://www.motorolasolutions.com/es_xl/productos/tetra.html) (14 marzo 2019)