



Memoria Justificativa y Solicitud de Contratación para contratos de RDLSE

OBJETO A CONTRATAR:

Suministro, Soporte y mantenimiento del software de la
plataforma Big Data Cloudera

NÚMERO DE LA S.C: 6000009932

Dirección: Dirección de Explotación
Ferroviaria. [Metro de Madrid, S.A.](#)

Área: Área de Comunicaciones
y Sistemas de
Información.

División: División de Instalaciones y
Sistemas de Información.

Servicio: Servicio de Integración de
Sistemas.

Aprobado por: Juan Tébar

1 OBJETO DE LA SOLICITUD DE CONTRATACIÓN

El presente documento tiene por objeto elevar a la aprobación del correspondiente órgano de contratación de Metro de Madrid, S.A., la autorización para el inicio de un proceso de licitación que tiene por objeto la contratación del suministro, soporte y mantenimiento del software de la plataforma Big Data Cloudera durante dos años, para los entornos de desarrollo y producción.

2 DATOS DE LA LICITACIÓN

▪ Objeto

Contratación consistente en el suministro, soporte y mantenimiento del software de la plataforma Big Data Cloudera durante dos años, de las siguientes licencias:

Entorno de Producción:

Tipo de licencia	Nº Suscripciones
Cloudera Enterprise Data Hub Edition, Node License, Gold Support	6 (Según tabla de configuración de nodos más adelante)
Cloudera Data Platform Flow Management Edition – Annual Subscription, 4 Cores - Business Support	16

Los servicios de soporte darán cobertura a un cluster de Big Data de 6 nodos, más dos nodos adicionales del producto Cloudera Data Flow, con las siguientes características:

Nodo	Cpu Cores	Memoria GB	Disco TB
Nodo Big Data 1	20	128	7
Nodo Big Data 2	20	128	7
Nodo Big Data 3	20	128	7
Nodo Big Data 4	28	384	31
Nodo Big Data 5	28	384	31
Nodo Big Data 6	28	384	31
Cloudera Data Flow 1	32	384	17
Cloudera Data Flow 2	32	384	17

Entorno de Desarrollo:

Tipo de licencia	Nº Suscripciones
Cloudera Enterprise Data Hub Edition, Node License, Bronze Support	4 (Según tabla de configuración de nodos más adelante)

Los servicios de soporte darán cobertura a un cluster de Big Data de 4 nodos con las siguientes características:

Nodo	CPU Cores	Memoria GB	Disco TB
Nodo 1	16	128	1
Nodo 2	16	128	1
Nodo 3	16	128	1
Nodo 4	16	128	1

▪ **Estamento responsable de la ejecución del contrato**

Servicio de Integración de sistemas

▪ **Valor estimado del contrato (Artículo 4 RDLSE)**

Valor estimado: 615.000,00 euros (IVA no incluido)

▪ **Método de cálculo aplicado para determinar el valor estimado (Art. 4 RDLSE)**

☒ El valor real de los distintos contratos análogos adjudicados durante el ejercicio precedente, ajustado en función de los precios del mercado sin incluir el IVA, teniendo en cuenta las posibles prórrogas (2 prórrogas de 6 meses).

▪ **Presupuesto base de Licitación (Art. 43 RDLSE)**

- Base imponible (BI): 410.000,00 euros
- Importe del I.V.A.: 86.100,00 euros
- Presupuesto base de licitación (PBL): 496.100,00 euros, IVA incluido

▪ **Desglose del presupuesto base de licitación (Art. 43.5 RDLSE)**

COSTE ENDÓGENOS Son los generados por el contrato y están formados por los Costes Directos e Indirectos.	
Costes Directos (98% del PE)	349.391,30 €
Costes Indirectos (2% del PE)	7.130,44 €
Presupuesto de Ejecución (PE): Costes Directos + Costes Indirectos	356.521.74 €
COSTE EXÓGENOS Son los relacionados con las actividades de la empresa y están formados por los gastos de estructura (GG+BI).	
Gastos Generales (9% PE)	32.086,96 €
Beneficio Industrial (6% PE)	21.391,30 €
Base imponible	410.000,00 €
Presupuesto Base de Licitación (Base imponible + IVA)	496.100,00 €

▪ **Modificación del contrato (artículo 110 RDLSE)**

☒ No procede

▪ **División en lotes:**

☒ **NO se divide en lotes (Art. 52.3 RDLSE)**

- **Justificar los motivos** de la no división en lotes: Este contrato se realiza para el mantenimiento de un grupo conjunto de licencias, servicio que proporciona el fabricante del mismo con un contrato global de licenciamiento. Desde el punto de vista técnico no es divisible en lotes porque el servicio global afecta a todas las licencias que se posean del fabricante.

▪ **Duración del contrato**

- Plazo de duración/ejecución inicial del contrato: 24 meses, a partir del día siguiente a la firma del acta de inicio de los trabajos o en la fecha de inicio que se indique en la propia acta, no comenzando antes del 1 de septiembre de 2022.
- Prórrogas:

☒ Sí

- N° de prórrogas: 2
- Duración de cada prórroga: 6 meses
- **Justificación** de la necesidad de prórrogas: Al tratarse de suscripción de licencias y mantenimiento, no puede haber intervalos de fechas entre el contrato que finaliza y el nuevo contrato porque no se podrían utilizar los productos o porque analizado el mercado se considera oportuno por el órgano de contratación.

▪ **Clasificación del contrato**

Sujeto a RDLSE (Real Decreto-ley 3/2020)

▪ **Naturaleza del contrato**

☒ Mixto (Suministros)

▪ **Procedimiento de licitación**

☒ Procedimiento Abierto

- **Justificación del procedimiento:**

No se reúnen los requisitos necesarios que exige el RDLSE para la aplicación de un procedimiento negociado. Por lo anterior, y con el fin de asegurar los principios del artículo 27 del RDLSE, se propone la contratación mediante procedimiento abierto.

▪ **Criterio de adjudicación (Art. 66 RDLSE)**

☒ Único criterio (precio o criterio basado en rentabilidad)

- **Justificar las razones** por el que se propone este criterio de adjudicación:
Al ser un suministro de licencias de Cloudera y el inseparable soporte asociado a las mismas, cuyas prestaciones vienen definidas por el fabricante de software Cloudera, no es posible introducir modificaciones de ninguna clase en el contrato.

▪ **Subcontratación (artículo 107 RDLSE)**

☒ Procede

- Indicar las tareas críticas que no podrán ser objeto de subcontratación: No hay tareas críticas

▪ **Procedimiento de subasta electrónica o petición sucesiva de ofertas**

☒ NO

▪ **Fondos FEDER**

☒ Contrato no financiable con fondos FEDER

▪ **Confidencialidad de los Pliegos de Prescripciones Técnicas**

☒ NO

▪ **Cesión de datos**

¿La ejecución de este contrato requiere la cesión de datos por parte de Metro de Madrid, S.A. al contratista?

☒ NO

3 JUSTIFICACIÓN DE LA NECESIDAD

Metro de Madrid dispone de una ingente cantidad de información basada en datos de distinta naturaleza procedente de diversos ámbitos de la compañía. Gran parte de estos datos se almacenan y procesan en sistemas propietarios de forma aislada por cada área funcional de negocio y en ciertos casos se eliminan pasado un tiempo para evitar la sobrecarga de los sistemas operacionales, la eliminación de estos datos supone una pérdida de información relevante que podría ser utilizada en estudios de análisis de datos.

El tratamiento y análisis de los datos de la compañía para poder extraer información de los mismos requiere en muchos casos de tareas de preprocesamiento de cara a mejorar su calidad o a adaptarlos según las necesidades de análisis específicas. Estas tareas requieren de grandes capacidades de procesamiento y almacenamiento y herramientas específicas en realizar este tipo de tratamiento de los datos.

Otro aspecto importante en el tratamiento de la información, es el análisis online de los datos que se generan en los distintos ámbitos de la compañía con la operativa diaria, y cuyo análisis en tiempo real representa un valor añadido importante para las distintas áreas de negocio ya que aporta una visión actual del estado operativo.

Así pues, el objetivo de este sistema es el de disponer del mantenimiento de la infraestructura de una plataforma de Big Data de Metro de Madrid. La razón de tener en nuestra instalación una plataforma así es la dotar a Metro de Madrid de una herramienta en la ayuda de toma de decisiones, basado en la recolección de enormes cantidades de datos heterogéneos de instalaciones, material móvil y cualquier otro que se precise para realizar informes, análisis, predicciones e incluso obtener conclusiones basadas en inteligencia artificial de un modo mucho más ágil que los existentes hasta ahora. Todo ello enfocado a la mejora del transporte ferroviario y al servicio que Metro presta.

Para poder mejorar estas capacidades de análisis, gestión, limpieza y procesamiento de toda la información de la compañía, así como poder implementar procesos de análisis de datos en tiempo real, de forma global y unificada, METRO se dotó en el año 2017, tras realizar un estudio de distintas herramientas de mercado (SC 2000002196), de la plataforma de Big Data Cloudera.

Esta plataforma permite el procesamiento masivo y distribuido de todos los datos procedentes de diversas fuentes de información de distinta naturaleza, conformando un repositorio único y central de información que podría ser alimentado por distintas herramientas y procesos de operación de la compañía, y proporciona las herramientas

necesarias para construir una arquitectura de procesamiento de los datos de forma online.

Este gran repositorio de procesamiento y almacenamiento de datos, tanto estructurados como no estructurados, permite también la integración de mecanismos y componentes para la explotación de la información almacenada de forma eficiente, tanto en escenarios de tiempo real como en escenarios de análisis batch de la información, ofreciendo resultados con unos tiempos de respuesta que se adapten a las necesidades actuales de METRO.

Desde el año 2017, mediante trabajos internos y con la ayuda de soporte técnicos especializados, de forma paulatina, se han ido implementando distintos casos de uso y se han ido incorporando nuevas fuentes de datos en la plataforma de Big Data. Esto ha permitido la incorporación a la plataforma de usuarios que anteriormente accedían a los sistemas operacionales de la compañía para realizar distintos tipos de análisis de datos, de forma que se ha descargado de carga de trabajo de dichos sistemas, mejorándose el rendimiento y permitiendo a los usuarios realizar análisis de datos que hasta ahora tenían limitados, ya que la plataforma de Big Data les ofrece muchas más capacidades.

La plataforma Big Data Cloudera se compone de distintos tipos de componentes técnicos, basados la mayoría en proyectos open source, que se utilizan para distintas tareas. Por ejemplo, la plataforma dispone de distintos componentes orientados a la carga y procesamiento de datos tanto en escenarios de tiempo real como en escenarios de análisis batch en función de cada caso de uso funcional.

Adicionalmente a los productos open source adaptados a un ecosistema integrado, el fabricante ha desarrollado componentes específicos que facilitan la gestión, el desarrollo de aplicaciones y el mantenimiento de toda la plataforma.

La unión del licenciamiento de los componentes específicos más el soporte y mantenimiento de todo el ecosistema es lo que da lugar a la necesidad de esta contratación.

El modo de licenciamiento de la plataforma se basa en suscripciones anuales en base al número de nodos que conforman el clúster de la plataforma. Estas suscripciones dan derecho al uso y soporte especializado del fabricante para cada producto incluido en la plataforma, tanto para la resolución de problemas como para el asesoramiento técnico y funcional en el uso de los mismos.

Dada la especialización técnica necesaria para dar soporte a este tipo de herramientas, que requieren de un conocimiento que actualmente no dispone Metro de Madrid de forma interna, para poder garantizar un nivel de servicio adecuado ante posibles incidencias y poder utilizar las herramientas específicas del fabricante para la gestión de la plataforma, se considera imprescindible contar con el soporte del fabricante del producto mediante la adquisición de las suscripciones de las licencias, y poder garantizar el funcionamiento adecuado de la plataforma en los casos de uso ya implantados e ir ampliando el número de usuarios que utilizan la plataforma.

4 ANTECEDENTES

	CONTRATACIONES ANTERIORES	CONTRATACIÓN ACTUAL
SOLICITUD DE CONTRATACIÓN N° CONTRATO	6000008403 7220000581
OBJETO DEL CONTRATO	Soporte y mantenimiento de licencias Cloudera.	Mantenimiento y soporte de las licencias Big Data Cloudera, entornos de desarrollo y producción.
DURACIÓN INICIAL DEL CONTRATO	24 Meses	24 Meses
PRÓRROGAS PREVISTAS	Si	Si
MODIFICADOS PREVISTOS	No	No
LOTES	No	No
PRESUPUESTO BASE LICITACIÓN (SIN IVA)	150.000	410.000
VALOR ESTIMADO	225.000	615.000

No existe comparación homogénea entre las contrataciones anteriores y la presente debido principalmente a dimensiones de plataforma, elementos licenciados y cambios en los modelos de licenciamiento.

Todo ello queda razonado a continuación.

En el año 2018, en el contrato n° 7718000050 se contrataron 6 suscripciones de tipo Cloudera Enterprise Data Engineering con el tipo de soporte Bronze, que daba derecho a un número limitado de componentes de la plataforma en horario laboral. El importe fue de 37.249,00 € porque se incluían también en la misma contratación los servicios de consultoría necesario para la instalación y securización de la plataforma, pero el precio de las suscripciones fue de 13.536,00 €.

En el año 2019, en el contrato n° 7219000264 se contrataron 6 suscripciones de tipo Cloudera Enterprise Data Engineering con el tipo de soporte Bronze, por un importe de 13.536,00 €. Durante el año 2019 la empresa Cloudera se fusionó con la empresa de la competencia Hortonworks e incluyó en su portfolio de productos los productos de ambas empresas. Esto llevó a la empresa Cloudera en octubre de 2019 a modificar la política de licenciamiento que estaba vigente hasta la fecha para las nuevas suscripciones que se realizasen a partir del año 2020, lo que supuso en su momento un incremento del coste de las licencias.

También ese incremento se debía a en el contrato 7220000581 se contrataron 6 suscripciones de tipo Cloudera Data Hub y el Soporte de la plataforma Cloudera Data Flow siguiendo el nuevo modelo de licenciamiento. También se incluía soporte de tipo Gold, que da derecho a un soporte 24x7, ya que el fabricante del producto lo había

ofrecido a Metro de Madrid sin incremento de coste dado el impacto de subida de precios por el cambio de modelo.

Con respecto a la nueva licitación, cabe reseñarse que hasta la contratación anterior el fabricante permitía realizar varios tipos de suscripciones que se diferenciaban principalmente en el número de componentes de la plataforma que se incluían, de forma que se podía licenciar solo algunos componentes. Metro de Madrid hasta ese momento estaba utilizando este tipo de suscripciones ya que, por los casos de uso que se tenían implementados, no era necesario utilizar el resto de componentes.

Con el nuevo modelo de licenciamiento, el principal cambio es que **ha eliminado el tipo de soporte limitado de componentes** y solo hay posibilidad de un único tipo de soporte general por nodo que incluye casi todos los componentes. Este tipo de licencia era análogo al tipo de suscripción completa del modelo anterior que era el que tenía mayor coste. Pero, como veremos a continuación, aunque no hubiese ocurrido ese cambio, el impacto en coste hubiera sido similar.

Otro cambio que incluye el nuevo modelo de licenciamiento es que **algunos componentes han sido extraídos del soporte general y ahora se licencian con suscripciones aparte para ese producto**. Este es el caso de los componentes que permiten el procesamiento de los datos en tiempo real que en el nuevo modelo se licencian con su propia suscripción. Este componente lo está utilizando Metro en prácticamente todos los casos de uso implementados.

Por último, se ha añadido la obligatoriedad de licenciamiento por tamaño de la plataforma de forma análoga a como hacen otros fabricantes. Se debe licenciar la plataforma en función del número de CPUs, capacidad de almacenamiento y memoria de la infraestructura.

Así pues, todos estos cambios de escenario han provocado este incremento significativo de coste del contrato con respecto a las anteriores contrataciones.

Entrando en más detalle de uno de esos puntos, cabe decirse que, como ya quedó enunciado, aunque con la nueva política de soporte del fabricante no es posible elegir ya entre el soporte limitado de componentes o el soporte completo de todos los componentes, Metro de Madrid ya se planteó la necesidad de cambiar el tipo de suscripción al soporte completo de componentes ya que el soporte que se contrató inicialmente, si bien daba cobertura a las necesidades iniciales de Metro, con la incorporación de nuevos usuarios y nuevos casos de uso a la plataforma, los departamentos necesitaban utilizar una serie de componentes que no estaban cubiertos en el soporte, por lo que era necesario contratarlos para poder garantizar el servicio, su uso legal a estos usuarios y soporte ante posibles fallos o errores en la plataforma.

Esto significa que, aunque esa empresa no hubiese cambiado el modo de licenciamiento, por la dimensión que esta plataforma está tomando, ya Metro se había planteado llegar a un modelo de soporte equivalente al que ahora es obligado hacer, tanto en tamaño de plataforma como en número de componentes.

En la contratación actual se introduce también el licenciamiento de la plataforma de desarrollo que Cloudera definió como obligatorio a partir del 2020. Durante los años 2020 y 2021 el licenciamiento del entorno de desarrollo se realizó mediante las

solicitudes de contratación de contratos menores 6000008950 y 6000009468. Este último contrato finaliza el 01/07/2022. Se realizará un nuevo contrato menor para cubrir los dos meses de soporte del entorno de desarrollo necesarios hasta el comienzo del que se propone en esta SolCon con los dos entornos conjuntamente.

En la contratación actual también se ha ampliado el número de cores licenciados para el componente Cloudera Data Flow, componente especializado para el procesamiento de datos online, para poder dar cobertura a todos los proyectos de ingesta y procesamiento en tiempo real, tales como el Tren Digital o el Cálculo del Grado de Ocupación Online.

Debido al aumento del número de proyectos que demandan el uso de esta tecnología y las necesidades de procesamiento de los mismos, también se hace necesario aumentar la configuración de hardware del entorno de desarrollo, lo que supone también un incremento en el importe para esta contratación.

5 JUSTIFICACIÓN

Todo este escenario de costes deriva de la evolución de la arquitectura hardware y de las licencias que dan soporte a la plataforma de Big Data que a su vez es una consecuencia de la ampliación de los casos de uso implementados, así como del número de usuarios y departamentos que hacen uso de la misma y a los que se está prestando servicios en sus diferentes necesidades funcionales.

Dada la complejidad de estas arquitecturas y la poca madurez de Metro en el uso de las mismas, en las fases iniciales de la implantación de la plataforma se abordaron de forma inicial varios prototipos de alcance limitado que permitieron a Metro ir sentando las bases para futuros proyectos y posteriormente se han implementado proyectos que aportan gran valor a la organización y con los cuales se ha dado un salto cuántico en el manejo de información de cara a la mejora de la prestación del servicio y de la optimización de costes en los mantenimientos.

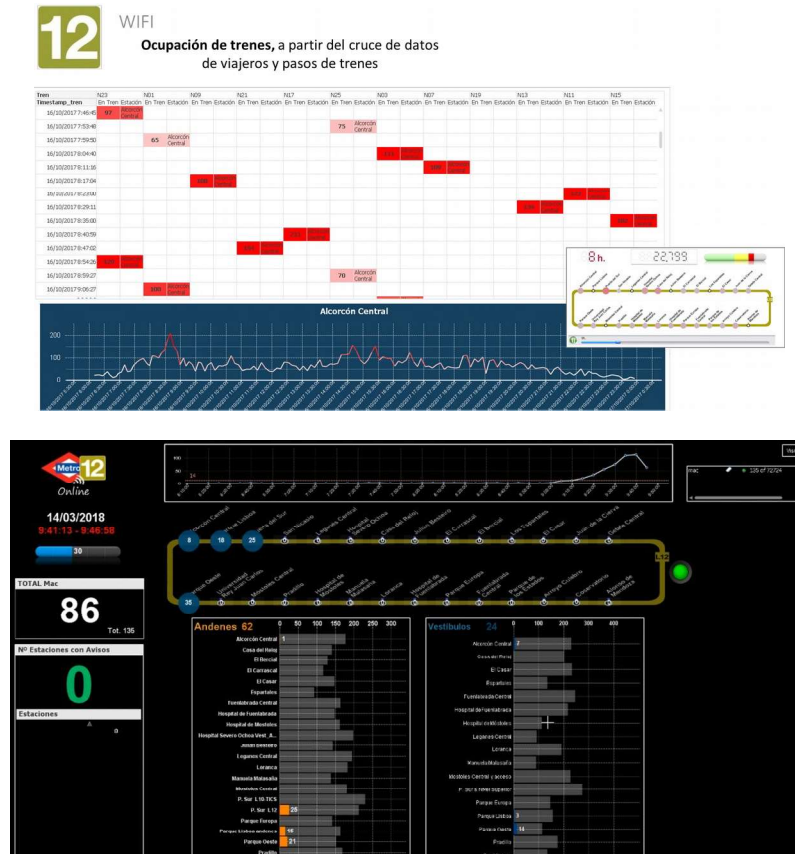
Así, pues, como ejemplo de la aportación de estas nuevas tecnologías a la empresa, se hace a continuación un repaso de los siguientes usos de la misma:

Datos WIFI L12

Durante el año 2018 junto con la instalación de la plataforma se implementó un prototipo relacionado con información relativa al posicionamiento del cliente, se basaba en la captura de las señales wifi de los dispositivos móviles de los viajeros que permitía por un lado identificar el volumen de viajeros en un vestíbulo o estación determinado y por otro correlacionar la identificación de las señales consecutivas de

los dispositivos identificados, para construir los trayectos entre estaciones realizados por los viajeros y estimar la ocupación de los trenes en dichos trayectos.

El prototipo se circunscribió a los datos recopilados de línea 12 por limitación de la infraestructura wifi disponible en ese momento.



Cálculo de velocidades CTC

Durante el año 2019 se implementó el primero de los proyectos. Se basaba en la captura y procesamiento de los eventos de pisadas de trenes a partir de la información suministrada por el sistema CTC.

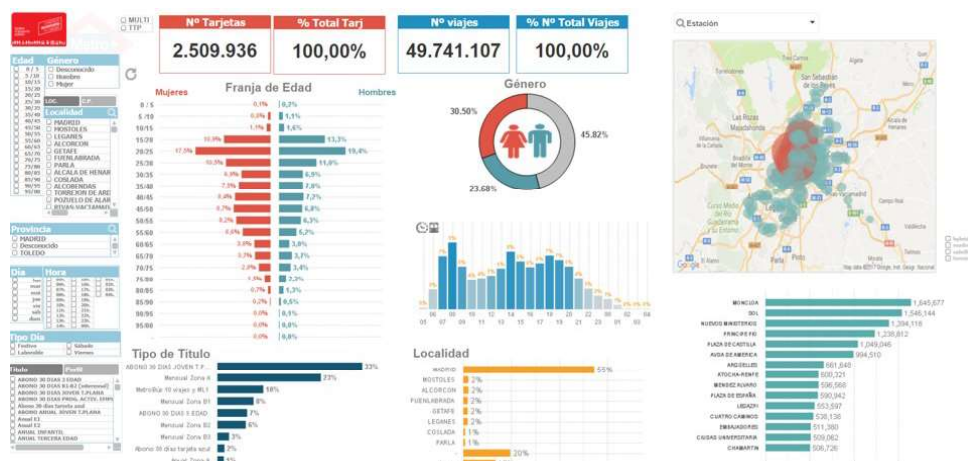
Este proyecto surgió a raíz de la necesidad de suministrar al servicio de Vía una base de datos que recopilara información de la velocidad media del paso de los trenes por los circuitos de vía, se realizó un desarrollo sobre la plataforma de big data que ingestaba y procesaba de forma online todos los eventos suministrados por el sistema CTC, correlacionando las pisadas y liberaciones de cada circuito de vía para calcular la velocidad media del tren.

Este proyecto permitía por un lado consultar la información procesada de forma online y construir una base de datos históricos de pisadas y velocidades sobre la que se

Aprovechando la arquitectura y el desarrollo realizado en la plataforma de big data para la información de pisadas de trenes del sistema CTC, se amplió la información que se capturaba y procesaba, para incluir no solo las ocupaciones y liberaciones de los circuitos de vía sino también el estado de los distintos elementos de la vía, como señales o agujas, y los mensajes de seguridad emitidos por el sistema CTC.

Caracterización de Viajeros

Uno de estos casos es la carga de los datos que suministraba el Consorcio Regional de Transportes sobre caracterización de clientes, estos datos se cargan en la plataforma para que puedan ser cruzados con otros de otros departamentos y accedidos desde distintas aplicaciones.



Matriz Origen Destino

En el 2019, con el fin de poder disponer de mayor información sobre la movilidad de los clientes de metro para poder planificar mejor el servicio, se llevó a cabo el proyecto de implantación de la generación de la Matriz Origen Destino, en este proyecto se implementó un algoritmo que en base a todas las entradas diarias de viajeros en la red, estima la salida más probable y genera una matriz con estación origen y estación destino de toda la red, lo cual permite modelar el flujo de viajeros de Metro.

Hasta la implantación de este proyecto en la plataforma de Big Data, Metro solo disponía de una información similar basada en una encuesta que se realizaba cada varios años debido a su alto coste. Los datos de la encuesta, además de ser estáticos y muy agrupados, no eran muy ricos; ya que no se distinguía en la encuesta los distintos tipos de día, ni las franjas horarias, ni estacionalidades, por lo que el resultado era una única matriz de un día tipo, realizada cada varios años.

A partir del proyecto de la Matriz Origen Destino, Metro empezó a disponer de forma diaria de una matriz origen destino cada 15 minutos, basada en entradas reales en la línea. El salto en calidad de información que empezó a utilizarse para la planificación del servicio fue muy grande, ya que a partir de ese momento Metro disponía de información muy granular, donde podía modelar en función de tipos de día, de franjas horarias o incluso de información real de cómo afectaba a los flujos de viajeros, los cierres de estaciones o los cortes de línea. Se pasó de tener una matriz cada varios años con información de una encuesta, a tener 80 matrices diarias con información de entradas reales.

No fue posible llevar a cabo este proyecto hasta disponer de forma operativa de la plataforma de Big Data, ya que el volumen de información a procesar y los cálculos necesarios excedían de las arquitecturas hasta la fecha disponibles en Metro.

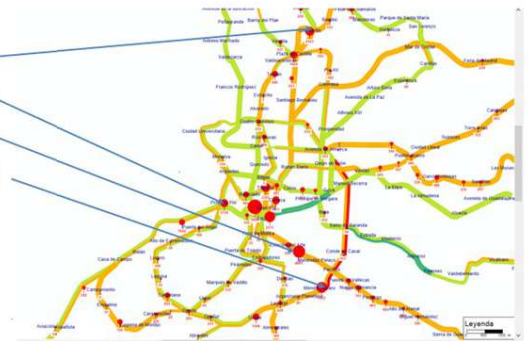
Este proyecto además de generar las matrices origen destino de forma agregada construye una base de datos con el histórico de movilidad de cada tarjeta de metro que permite realizar distintos análisis de patrones de comportamiento en función de los distintos viajes.

La información generada por este proyecto es utilizada en distintos departamentos. El Servicio de Planificación y Estudios de Operación lo utiliza como fuente de múltiples estudios y aplicaciones funcionales dentro de su ámbito, entre otros, en el estudio de la movilidad ante eventos o cambios de la movilidad ante modificaciones de otros medios de transporte, ante cambios en la red de Metro por cierres o ampliaciones.



Incrementos corte Cercanías

- Chamartín
- Príncipe Pío
- Atocha-Renfe
- Méndez Álvaro



Captación de cambios en la Red



Cambios de comportamiento debido a cortes en la Red

Ejemplo cierre estación de Sevilla L-2



Esta información también es utilizada por el Servicio de COMMIT para conocer la afluencia de personas en ciertas estaciones y por el Servicio de Calidad como fuente para realizar ciertos estudios.

Aforos de Estaciones

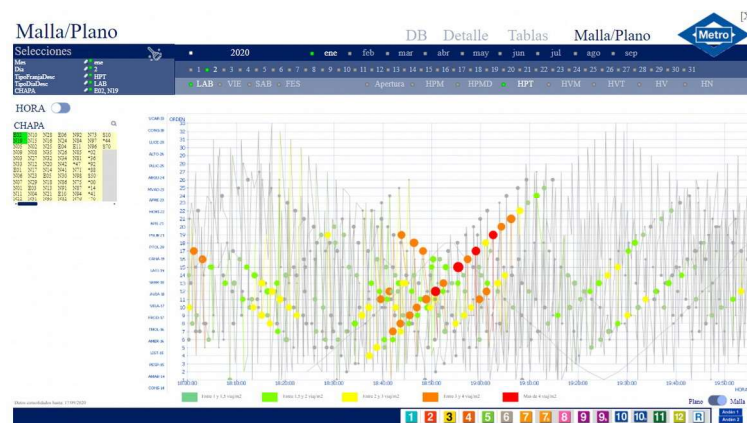
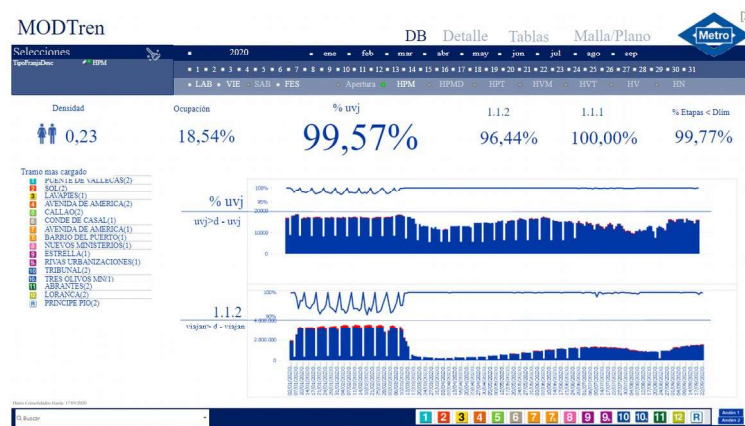
La plataforma de Big Data fue utilizada durante la pandemia por el Área de Operación para la implementación del control de aforos de estaciones, para generar la fuente de datos que estimaba el flujo de viajeros necesario para evitar aglomeraciones y así poder calcular el aforo máximo permitido. Se utilizó la plataforma de Big Data ya que era necesario realizar consultas y cálculos sobre un volumen de datos muy elevado.



Cálculo de grado de ocupación

Durante el año 2020 se implementó un proyecto que, basándose en la información que se generaba en las matrices origen destino relativa al flujo de viajeros del día anterior y la circulación de los trenes en la red del mismo día, implementaba un algoritmo que realizaba la correlación entre ambas fuentes de datos y estimaba la carga de viajeros en función de las entradas y del tren más cercano en el tiempo, dando lugar al cálculo del indicador del grado de ocupación de trenes.

Esta información sirve de fuente para los cuadros de mando de la operativa relativos al flujo de viajeros y el grado de ocupación.



Estadísticas Ocupación APP Móvil

Durante el 2021, Metro decidió publicar parte de esta información estadística para consulta de los viajeros en la aplicación móvil. Se pueden consultar días similares para saber cómo ha sido la circulación y poder planificar mejor su viaje.

Para llevar esto a cabo, se ha desarrollado un proyecto dentro de la plataforma que, en base al histórico de información del grado de ocupación, agrega los datos estadísticos, limpia los datos anómalos debidos a perturbaciones y publica diariamente en la aplicación móvil los datos para el viajero.



Análisis Calidad Scada Recaudación y Venta

Otro proyecto que está en curso en las fases finales es el de validación de datos de recaudación y venta. El objeto de este proyecto es implementar un motor de reglas configurable que permita cargar datos de venta y realizar validaciones de los mismos de forma automática sin que sea necesaria la intervención de los usuarios funcionales. Esto permitirá validar de forma diaria los datos de recaudación y venta y generar un informe con las incidencias detectadas.

Hasta el año 2020, la mayor parte de los casos de uso que se implementan en la plataforma de Big Data son funcionalidades que requieren grandes capacidades de procesamiento masivo de la información en Batch. En este ámbito estarían los proyectos mencionados sobre el cálculo de la matriz origen destino, el cálculo del

grado de ocupación de trenes o las distintas capacidades analíticas que utilizan los diferentes departamentos para explorar y analizar la información histórica almacenada en el repositorio de información de Big Data.

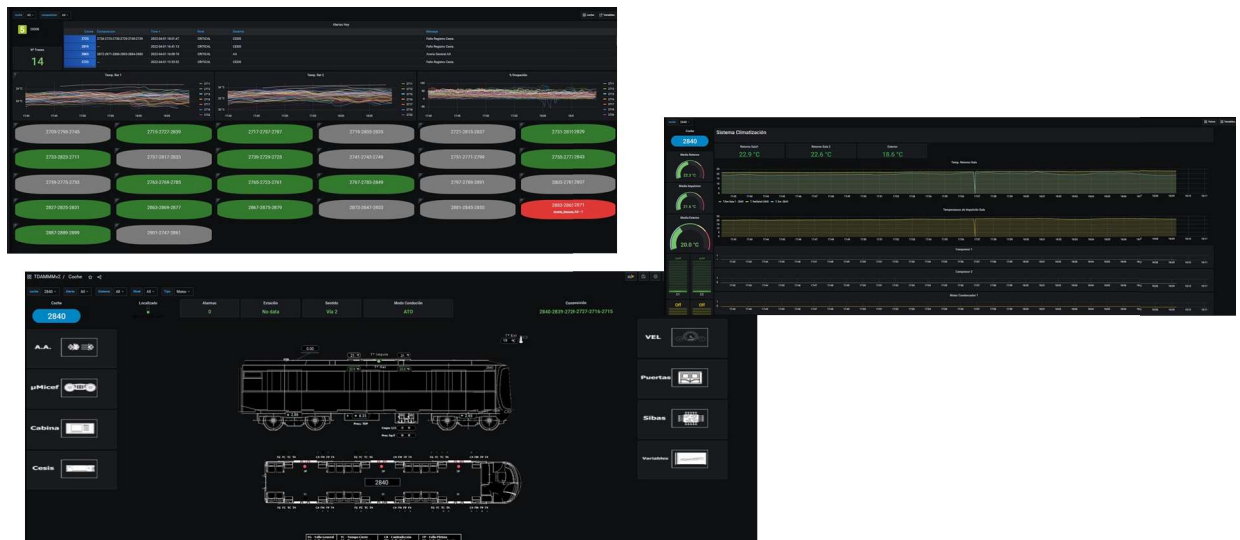
Tren Digital

En el año 2020 se amplió la infraestructura hardware y los componentes software de la plataforma de Big Data para dotarla de las funcionalidades necesarias de procesamiento en tiempo real de la información, que hasta ese momento estaban limitadas. Esto permitió acometer nuevos proyectos para dar solución a casos de uso cuya problemática principal era la generación continua de datos que era necesario capturar, procesar y almacenar.

Dentro de este ámbito funcional se implementaron diversos proyectos entre los que cabe destacar los proyectos relacionados con el tren digital realizados junto con el Área de Mantenimiento de Material Móvil. Desde el año 2020, este área diseñó un dispositivo embarcado que recolectaba información de los trenes 2000B de línea 5 y 2000A de línea 1. Esta información debe ser enviada a una plataforma de tierra que fuera capaz de recolectar la información, procesarla y almacenarla para poder ser analizada por el personal de mantenimiento.

Utilizando los nuevos componentes adquiridos en la plataforma de Big Data para el procesamiento de datos en tiempo real, se llevaron a cabo varios proyectos que permitieron implementar las diferentes funcionalidades necesarias para el Área de Mantenimiento de Material Móvil.

Actualmente mediante estas herramientas, el AMMM puede monitorizar el funcionamiento en tiempo real de la flota de trenes 2000, tanto en línea 1, como en línea 5, pudiendo detectar anomalías en el funcionamiento de los trenes antes de que provoquen una avería en línea.



Estas funcionalidades también permiten obtener información online del tren en un dispositivo móvil desde su trabajo de campo diario.



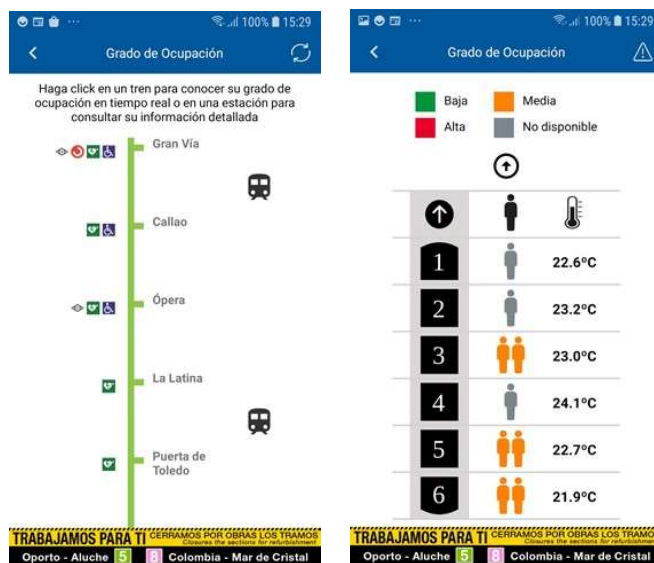
Durante el año 2021 se trabajó también con el Área de Ingeniería de Material Móvil para implementar a través de la plataforma de Big Data funcionalidades similares del tren digital para la flota de trenes 3000 de Línea 5.



Grado de Ocupación L5 App Móvil

Con la monitorización de los trenes 3000 de línea 5, quedaba completamente monitorizada toda la flota de trenes de dicha línea. Esto ha permitido realizar un proyecto que actualmente se encuentra en sus fases finales de validación que es el de la publicación de los datos de ocupación de los trenes de línea 5 en la app móvil de metro de Madrid para que pueda ser consultada por los viajeros de forma online.

Este proyecto se basa en la plataforma de Big Data y sus herramientas de procesamiento online de la información que capturan la información que reciben de los trenes 2000 y 3000 de forma online, la cruzan con los datos recibidos también de forma online desde el sistema CTC y GDL para localizar en línea el tren y asignarle una chapa y dan formato a la información para que pueda ser consumida por la aplicación móvil.



Actualmente también se está estudiando diseñar un sistema similar al implementado para los trenes 2000 de línea 5 para los trenes 6000 de línea 9B.

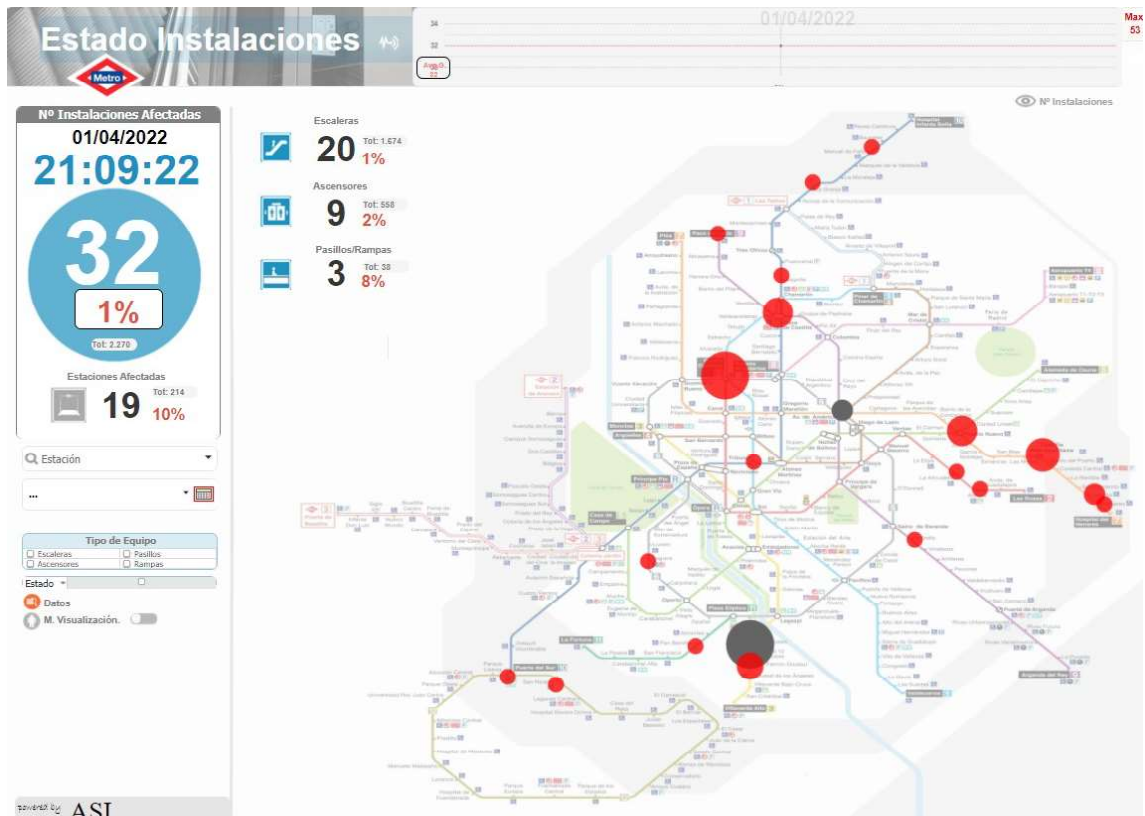
Temperaturas Estación

Otro proyecto que se está llevando a cabo en este momento, relacionado con la captura de información en tiempo real, es el proyecto para publicar los datos de las sondas que recogen la temperatura de las estaciones y cuartos técnicos y ponerlos a disposición del personal del Área de Mantenimiento de Instalaciones. Toda la arquitectura para la captura de datos desde el sistema SCADA de instalaciones, el procesamiento y almacenamiento de los mismos se realiza a través de la plataforma de Big Data. Esto permitirá, tanto publicar la información online, como construir un repositorio de información de datos históricos para realizar distintos análisis de datos que hasta ahora se estaban desechando.

Por otra parte, se está trabajando con el Servicio de Vía para diseñar un sistema similar para la captura online de información desde los engrasadores de vía, que permitirá al personal de vía monitorizar los distintos engrasadores, así como analizar la información reportada. Este proyecto se encuentra actualmente en fase de diseño.

Estado Instalaciones

Durante el año 2018 se implementó una funcionalidad que permitía monitorizar el estado de las instalaciones en tiempo real y visualizarlo a través de un cuadro de mando de forma agregada. Dado que el volumen de estados era muy grande, solo se capturaba información cada 10-15 minutos y la información no estaba modelada sobre una base de datos por lo que era difícil realizar análisis de la misma mas allá de los datos que mostraba el cuadro de mando. A partir de tener disponible la plataforma de Big Data, la arquitectura de captura y almacenamiento de datos se migró a la misma, lo que permitió por un lado una captura de datos de forma mas continua que permitiera disponer de información mas granular y, por otro, que los datos fueran accesibles y analizables a través de un repositorio de información que permitiera realizar con los mismos datos diferentes análisis a diferentes departamentos como el Servicio de Calidad o el Área de Mantenimiento de Instalaciones, además de poder cruzar la información del estado de las instalaciones con otros conjuntos de datos disponibles en la plataforma para poder enriquecer dichos análisis.



Repositorios Análisis Departamentales

Un problema recurrente que tienen muchos departamentos es que utilizan los sistemas operacionales para cubrir las necesidades de análisis de datos. Estos sistemas están orientados a dar servicio operativo para las transacciones de negocio, pero tienen limitadas las capacidades de análisis de datos, por lo que los usuarios muchas veces se ven limitados en los análisis que pueden realizar sobre estos sistemas. Dadas las altas capacidades de procesamiento, almacenamiento y análisis de información de la plataforma de Big Data de forma ágil y centralizada, un caso de uso general que se está implementando de forma incremental y que está dando servicio a distintos departamentos es el de modelar repositorios de información cargando los datos desde los sistemas operacionales a la plataforma centralizada de Big Data, para que los usuarios puedan explotar la información desde esta plataforma. Esto dota a los usuarios de unas capacidades de análisis que les permiten realizar estudios que hasta ahora no podían sobre los sistemas actuales. Además, se está construyendo una base de datos histórica con todos los datos que se están cargando que posibilita que puedan estar disponibles para distintos análisis, aunque sean limpiados en los sistemas fuente operacionales.

Estos repositorios se están implementando para dar servicio a distintos departamentos. Cabe destacar, entre otros, el servicio de COMMIT que está analizando desde la plataforma de Big Data los datos de alarmas y averías de equipos o los datos de los pozos de bomba que se están cargando desde el sistema de Scada de instalaciones.

Para el servicio de COMMIT también se ha implementado un repositorio con los datos del sistema SCADA de venta y peaje que sirve de fuente de datos para la aplicación KNIME con la que el servicio de COMMIT está desarrollando un modelo de mantenimiento predictivo para los datos de las incidencias de peaje.

Desde los distintos servicios de la división de operaciones se está dando servicio con distintos repositorios que almacenan datos de los diferentes servicios, por ejemplo, se están almacenando datos de peaje, de transacciones de venta, datos de intervención, de datos de coches kilómetro, etc.

Por todo lo comentado anteriormente, ha sido necesario ir haciendo crecer la plataforma de Big Data de forma horizontal, añadiendo más nodos, de forma vertical, aumentando la capacidad de cómputo de los nodos y añadiendo nuevos componentes software para ampliar las funcionalidades y poder ir dando servicio a la demanda creciente de los distintos departamentos funcionales.

Desde la implementación de los primeros prototipos y pruebas de concepto hasta la actualidad, la plataforma de Big Data ha ido creciendo en funcionalidades dando servicio a diferentes departamentos y sirviendo de fuente de datos para múltiples indicadores, funcionalidades y para información que se publica de cara al viajero.

La no renovación de las licencias de la plataforma de Big Data implicaría no poder utilizar legalmente dicha plataforma por lo que todas las funcionalidades descritas anteriormente no se podrían seguir utilizando en dicha plataforma. La sustitución de esas funcionalidades o la implementación de las mismas en otra arquitectura sería más o menos costosa dependiendo de la funcionalidad. Todas aquellos departamentos que utilizaran la plataforma como fuente de datos para análisis de datos deberían acceder a los sistemas operacionales fuente para obtener los datos. Esto podría no ser viable para ciertos tipos de análisis que necesitan unas capacidades de cómputo elevadas y sería necesario aumentar las licencias y los recursos hardware de los sistemas operacionales, con el coste económico que requiera, si se quiere seguir manteniendo el nivel de servicio que se daba a través de la plataforma de Big Data sin afectar al servicio transaccional que prestan los sistemas operacionales.

Las funcionalidades relativas a la matriz origen destino no se podrían utilizar hasta que se implementara en otro sistemas el desarrollo y los algoritmos desarrollados en la plataforma Big Data o volver al modelo anterior y generar las matrices en base a la realización de encuestas, lo que supondría un coste elevado si la encuesta se requiere que sea de amplio alcance y, como se ha comentado, se reduciría la calidad de los datos obtenidos, lo que redundaría en la calidad de los estudios que realiza la División Operativa para la planificación del servicio.

Las funcionalidades de cálculo del grado de ocupación y los indicadores que se calculan a través de los desarrollos implementados en la plataforma Big Data no se podrían obtener hasta la implementación en otro sistema de desarrollos similares, con el coste de implantación y licenciamiento del nuevo sistema. Esto afectaría también a las funcionalidades que publican datos estadísticos de ocupación de trenes a los viajeros en la app móvil que deberían dejar de publicarse hasta su implementación en otro sistema.

Todos los desarrollos relacionados con las funcionalidades del tren digital de línea 5 y línea 1 no se podrían utilizar hasta su desarrollo o implementación en otra plataforma con los costes de implantación y desarrollo que ello conlleva, o habría que dejar de utilizar los mismos, lo que redundaría en una peor calidad de servicio para el Área de Mantenimiento de Material Móvil, que no podría monitorizar los trenes para detectar averías de forma temprana. Tampoco se podría utilizar la información enviada por el tren digital para otras funcionalidades previstas como la publicación al viajero del grado de ocupación.

Y para los demás casos de uso resultarían consecuencias equivalentes que realmente son una vuelta al pasado del manejo de información basado en sistemas transaccionales que hoy en día no pueden responder a la demanda de información que requiere la empresa.

6 INFORMACIÓN PRESUPUESTARIA

PRESUPUESTO DE GASTO

AÑO	2022	2023
IMPORTE PERMITIDO	205.000€	205.000€
CECO	2601	2601
CUENTA	622226	622226

El presente documento, emitido a efectos de cumplimiento de obligaciones en materia de transparencia, es copia fiel del original, en el que constan las firmas auténticas y completas de las personas firmantes.

En cumplimiento de las obligaciones de protección de datos personales, no constan en esta copia datos identificativos adicionales a nombre y apellidos.