

Juan David Rosero Gómez A00400160  
Nicolás Jiménez Vargas A00400906  
Jose Daniel Guzmán Castrillón A00401303

## Ingeniería de software IV

Arquitectura de Software: Estimación de Velocidad Promedio por Tramos  
en Rutas con Datos del SITM-MIO  
<https://github.com/DanielGuzman13/FinalProjectIngesoft4-SITM-MIO.git>

### A. Informe de experimentos y ejecución

#### 1. Introducción y Objetivos

##### 1.1 Introducción

En el siguiente informe se presentan con la intención de mostrar los resultados del experimento de computación distribuida, donde se hace el procesamiento de la velocidad promedio de los arcos del caso SITM-MIO

##### 1.2. Objetivos del Experimento

El objetivo principal de este informe es evaluar la escalabilidad y la eficiencia de la arquitectura propuesta (punto C) para el procesamiento de datagramas del SITM-MIO, y determinar el punto de corte en el volumen de datos a partir del cual el uso de un sistema distribuido ( $N > 1$  nodo) es más eficiente que el procesamiento en un único nodo.

#### 2. Configuración Experimental:

La configuración experimental se realizó de la siguiente manera:

- Se tuvo en cuenta el uso de 4 nodos de procesamiento
- Se usaron desde 1 worker hasta 3 para la observación del cambio del performance.
- Se tuvo en cuenta el número de lotes que se iban a usar para el procesamiento distribuido (5.000, 10.000).
- Se tuvo en cuenta la cantidad de datos del datagrama history, en este caso se usaron 100.000 y 1'000.000

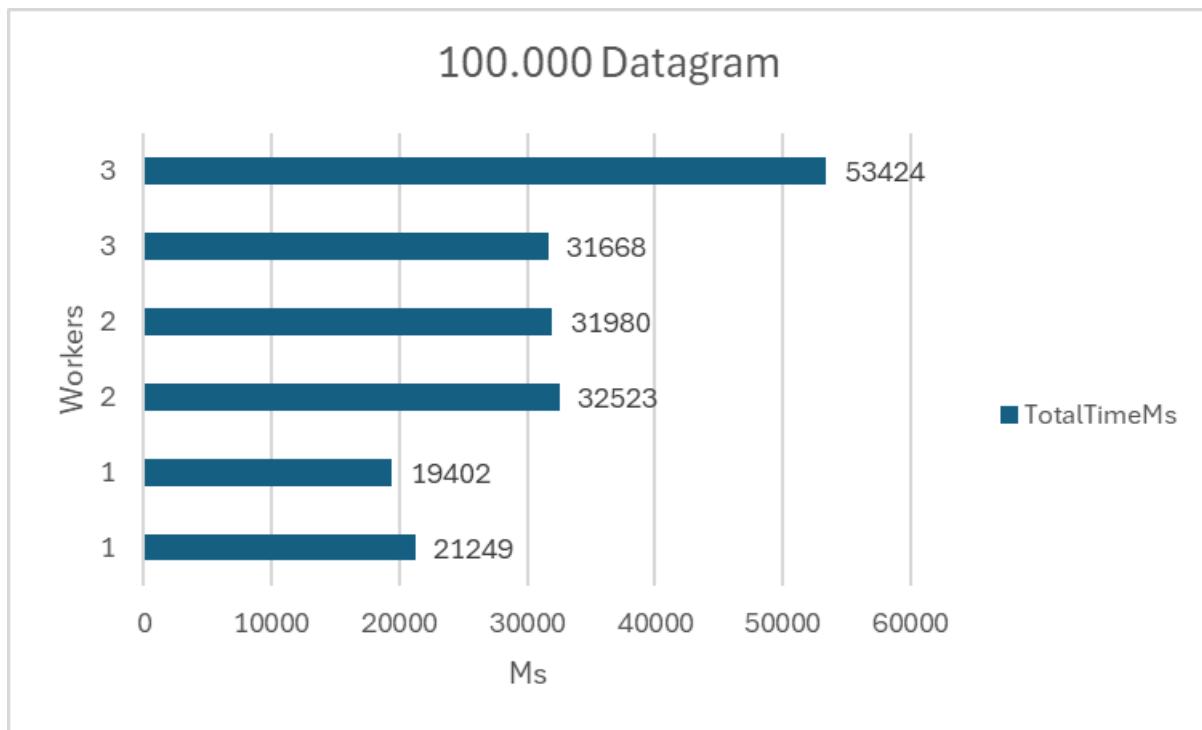
#### 3. Ejecución de protocolo:

El experimento se alternó entre las combinaciones de, la cantidad de worker, el número de lotes a usar y el tamaño de datos del datagram history.

Primero se aseguró de tener la caché del computador limpia, se cerraron todos los programas que podían consumir recursos, y antes de iniciar se reinicia el computador. Luego se inició la ejecución del experimento con los datos de tamaño 100.000 con 1 worker y 5.000 de lote, luego se aumentaba a 10.000 el lote, se repetía lo mismo aumentando los worker, para finalmente aumentar el tamaño de los datos a 1'000.000

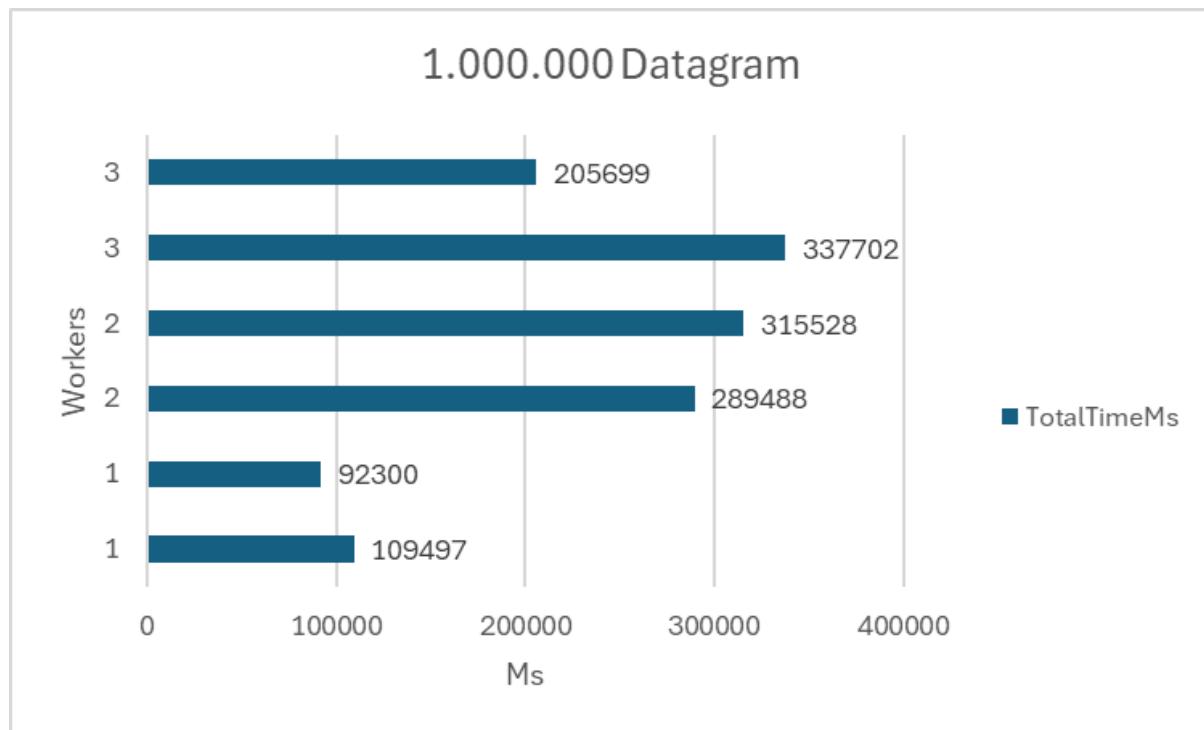
## 4. Resultados

Workers	TotalTimeMs	BatchSize
1	21249	5000
1	19402	10000
2	32523	5000
2	31980	10000
3	31668	5000
3	53424	10000



Juan David Rosero Gómez A00400160  
Nicolás Jiménez Vargas A00400906  
Jose Daniel Guzmán Castrillón A00401303

Workers	TotalTimeMs	BatchSize
1	109497	5000
1	92300	10000
2	289488	5000
2	315528	10000
3	337702	5000
3	205699	10000



## Conclusiones del Análisis de Escalabilidad y Eficiencia

1. Dominio de la Sobrecarga (Overhead) del Sistema Distribuido:
  - Para los volúmenes de datos pequeños y medianos probados (100.000 y 1.000.000 de datagramas), el Speedup ( $S_N$ ) fue consistentemente menor a 1.0 (los valores cayeron hasta 0.32).
2. Baja Eficiencia de los Nodos:
  - La Eficiencia ( $E_N$ ) cayó drásticamente al introducir más nodos, llegando a ser tan baja como 0.15 (15%) para el dataset de 1 millón de datagramas con  $N=3$ .
  - Esta baja eficiencia significa que los recursos adicionales (los Workers) no se están aprovechando bien, ya que la mayor parte del tiempo se dedica a la coordinación y no al procesamiento puro de datos.
3. Determinación del Punto de Corte:

Juan David Rosero Gómez A00400160  
Nicolás Jiménez Vargas A00400906  
Jose Daniel Guzmán Castrillón A00401303

- Se concluye que el punto de corte (el volumen mínimo de datos para el cual la arquitectura distribuida se justifica) se encuentra por encima de 1.000.000 de datagramas.

## B. Árbol de particionamiento refinado hasta el segundo nivel de profundidad

### ESPECIFICACIÓN DE SUB SISTEMAS

#### Sistema General: Centro de Control de Operación (CCO) – SITM-MIO

Unidad central encargada de gestionar la operación del SITM-MIO, procesar los datos enviados por los buses, analizar la movilidad urbana y ofrecer servicios de información en tiempo real y analítica a los controladores, administradores y ciudadanía.

#### 4. Subsistema de Persistencia y Procesamiento de Datos

Gestiona el almacenamiento, validación y análisis de grandes volúmenes de datos históricos en tiempo real.

- **4.1. Data Ingestion Layer (Capa de Ingesta):**

Recibe los datagramas de los buses (vía GPRS) y los normaliza.

(RF6)

- **4.2. Data Lake / Data Warehouse**

Almacena datos históricos y operativos para consultas analíticas y entrenamiento de modelos.

(RF6)

- **4.3. Módulo de Limpieza y Validación de Datos**

Filtra duplicados, errores y datos incompletos; garantiza consistencia temporal.

(RF18)

- **4.4. Módulo de Procesamiento en Streaming**

Procesa los datos entrantes en tiempo real sin interrumpir la disponibilidad del

Juan David Rosero Gómez A00400160  
Nicolás Jiménez Vargas A00400906  
Jose Daniel Guzmán Castrillón A00401303

sistema.

(RF11, RF12)

## 5. Subsistema de Análisis de Movilidad y Estimación de Velocidades

Analizar los datos históricos en tiempo real para calcular la velocidad promedio por arco y variables de movilidad.

- **5.1. Módulo de Cálculo Analítico**

Estima la velocidad promedio por arco y los tiempos de viaje promedio entre puntos.  
(RF7, RF19)

- **5.2. Motor de Actualización Dinámica**

Recalcula métricas en tiempo real ante la llegada de nuevos eventos o variaciones de tráfico.

(RF7, RF11)

- **5.3. Publicador de Resultados ( podemos considerar usar una API Interna)**

Expone resultados analíticos para ser consumidos por otros módulos o dashboards.  
(RF19)

## 8. Subsistema de Computador Embebido y Sensores del Bus

Captura datos físicos del vehículo y permitir la interacción del conductor con el sistema central.

- **8.1. Unidad de Sensores ( $\approx 40$  por bus)**

Capturan variables como velocidad, apertura de puertas, ubicación GPS, motor, etc.

- **8.2. Computador Embebido (On-board Computer)**

Procesa los datos de sensores y los empaqueta en datagramas enviados cada 30 s al DataCenter.

(Relación con RF1, RF2)

- **8.3. Interfaz de Usuario del Conductor (GUI/Perilla)**

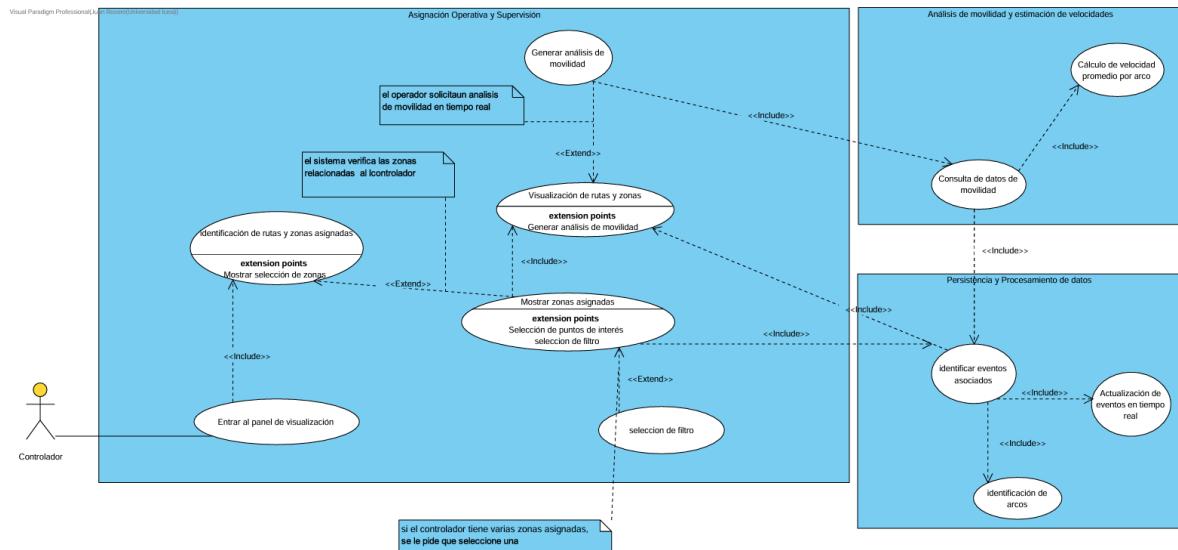
Permite reportar eventos manualmente con interacción mínima y segura.

(RF1, RF2)

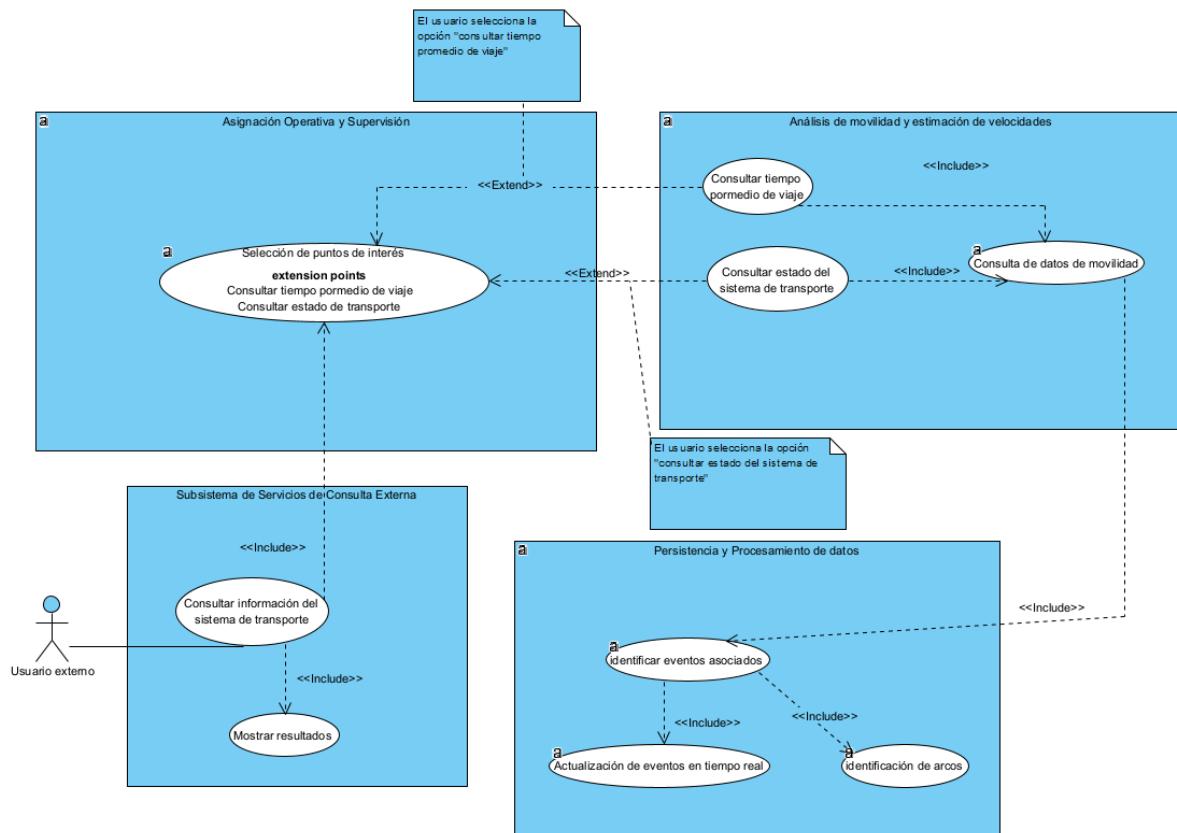
Juan David Rosero Gómez A00400160  
 Nicolás Jiménez Vargas A00400906  
 Jose Daniel Guzmán Castrillón A00401303

## C. Diagramas de casos de uso

### Caso de uso 1: Visualizar el análisis de movilidad por zona (R9 y relacionados)

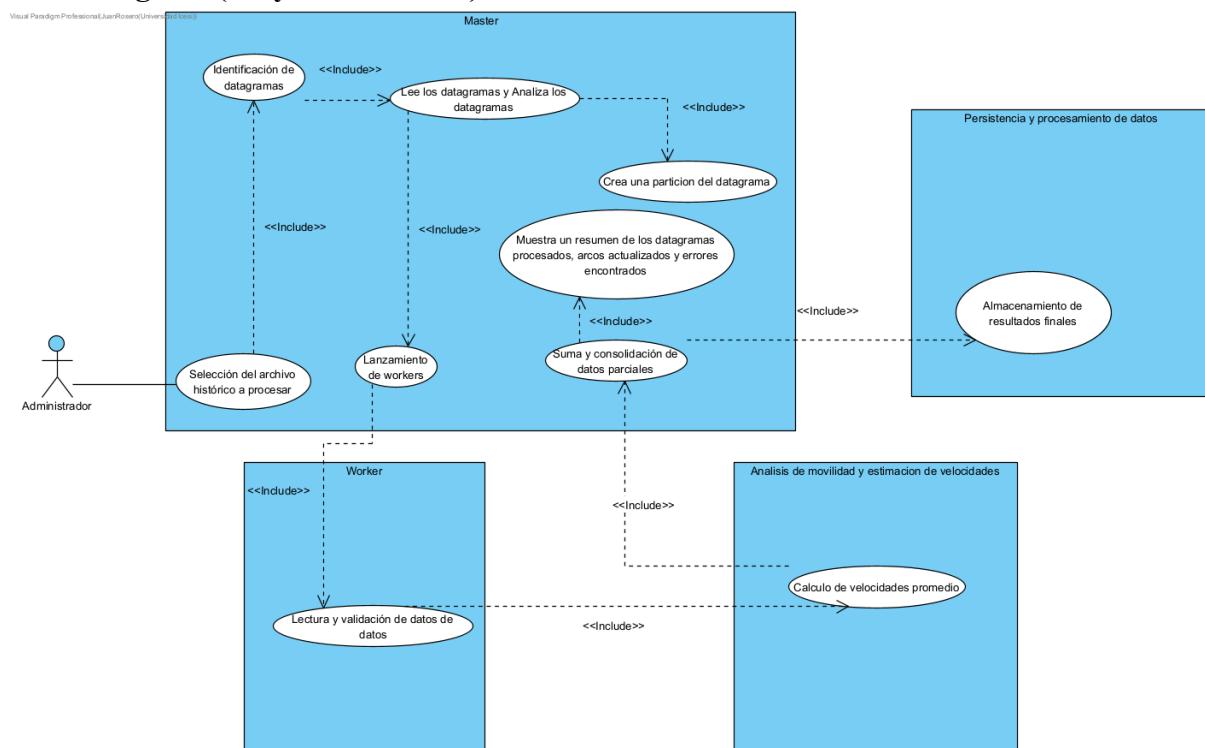


### Caso de uso 2: Consultar información del análisis generado (R13 y relacionados)



Juan David Rosero Gómez A00400160  
 Nicolás Jiménez Vargas A00400906  
 Jose Daniel Guzmán Castrillón A00401303

### Caso de uso 3: Calcular automáticamente la velocidad promedio histórica de todos los arcos del grafo (R6 y relacionados)



### D. Especificación de casos de uso en formato bicolumnar

- CU01RF09

<b>ID</b>	CU01RF09
<b>Autor</b>	Daniel Guzmán - Juan Rosero - Nicolás Jiménez
<b>Caso de uso</b>	Visualizar análisis de movilidad por zona
<b>Casos de uso relacionados</b>	CU01RF06, análisis de movilidad, asignación operativa, recepción y procesamiento de datagramas históricos.
<b>Actor</b>	Controlador
<b>Precondición</b>	El controlador está autenticado y ha iniciado sesión; la base de datos analítica contiene datos históricos procesados de movilidad (por ejemplo, velocidades promedio por arco); las rutas y zonas de la ciudad han sido asignadas previamente a los controladores de operación.
<b>Contexto</b>	Un controlador, ubicado en el CCO y responsable de una zona de la ciudad, desea visualizar indicadores de movilidad basados en los datos históricos de velocidad promedio por arco de su zona asignada.

Juan David Rosero Gómez A00400160  
 Nicolás Jiménez Vargas A00400906  
 Jose Daniel Guzmán Castrillón A00401303

<b>Acciones del usuario</b>	<b>Acciones del sistema (Incluye flujo normal y excepciones)</b>
1. El controlador ingresa al panel de visualización de rutas por zona asignada.	
2. El sistema identifica las rutas y zonas asignadas al controlador.	
	3 Si el controlador no tiene rutas y/o zonas asignadas, el sistema muestra un mensaje indicando que no es posible visualizar información de movilidad y finaliza el caso de uso.
	3.1 Si el controlador tiene más de una zona asignada, el sistema muestra la lista de zonas para que el controlador seleccione una.
4 El controlador selecciona una zona de la lista cuando tiene más de una zona asignada.	
4.1 Si el controlador solo tiene una zona asignada, continúa con la zona predeterminada sin necesidad de selección.	
	5 El sistema consulta en la base de datos analítica los datos históricos de movilidad de la zona seleccionada (por ejemplo, velocidades promedio por arco y otras métricas).
	5.1 Si no existen datos históricos para la zona seleccionada, el sistema indica que no hay información disponible y finaliza el caso de uso.
6 El controlador selecciona, mediante un filtro, una ruta o una categoría de evento	
	7. El sistema aplica el filtro por ruta o categoría de evento y actualiza la información mostrada.
8. El controlador selecciona, mediante otro filtro, un arco específico o un bus de interés	
	9. El sistema aplica el filtro por arco o bus y actualiza la información mostrada.
10 El controlador solicita ver los detalles del análisis para la selección realizada.	
	11 El sistema verifica si existen datos de movilidad para la combinación de filtros seleccionados.

Juan David Rosero Gómez A00400160  
 Nicolás Jiménez Vargas A00400906  
 Jose Daniel Guzmán Castrillón A00401303

	11.1 Si existen datos históricos, el sistema organiza la información y presenta las métricas calculadas, incluyendo la velocidad promedio histórica por arco y otros indicadores de movilidad.
	11.2 Si no existen datos históricos suficientes para los filtros aplicados, el sistema informa que no hay datos disponibles para esa combinación.
	12. El sistema muestra una representación visual de los resultados sobre el mapa, resaltando los arcos y elementos correspondientes y presentando las métricas de movilidad asociadas.
<b>Poscondición</b>	El controlador visualiza las métricas de movilidad basadas en velocidades promedio históricas por arco que corresponden a su zona y a los filtros aplicados.
<b>Observaciones:</b>	El caso de uso depende de los subsistemas de análisis de movilidad y de recepción y procesamiento de datagramas históricos.
<b>Observaciones:</b>	La información mostrada proviene de datos históricos previamente procesados; no se realiza procesamiento en streaming ni en tiempo real durante este caso de uso.

- CU01RF13

<b>ID</b>	CU01RF13
<b>Autor</b>	Daniel Guzmán - Juan Rosero - Nicolás Jiménez
<b>Caso de uso</b>	Consultar información del estado del sistema y estimaciones de tiempo promedio
<b>Casos de uso relacionados</b>	CU01RF09, CU01RF06
<b>Actor</b>	Usuario final
<b>Precondición</b>	La base de datos analítica contiene datos históricos procesados de velocidades promedio por arco y métricas de estado del sistema.
<b>Contexto</b>	Un usuario final (civil, empresa o entidad pública) accede al portal de consultas para visualizar el tiempo promedio de viaje entre dos puntos de la ciudad y/o el estado general del sistema de transporte, utilizando datos históricos procesados.

Juan David Rosero Gómez A00400160  
 Nicolás Jiménez Vargas A00400906  
 Jose Daniel Guzmán Castrillón A00401303

<b>Acciones del usuario</b>	<b>Acciones del sistema (Incluye flujo normal y excepciones)</b>
1. El usuario ingresa al portal de consulta.	
	2. El sistema muestra la interfaz de consulta con las opciones disponibles (por ejemplo: “Tiempo promedio de viaje” y “Estado del sistema de transporte”).
2.1 El usuario selecciona la opción “Tiempo promedio de viaje” e ingresa el punto de origen y el punto de destino.	
	2.1.1 El sistema verifica que los puntos de origen y destino sean válidos (existen paradas o zonas registradas en el grafo de rutas).
	2.1.2 Si los puntos no son válidos, el sistema muestra un mensaje de error indicando que no es posible calcular el tiempo de viaje y solicita corregir la información.
	2.1.3 Si los puntos son válidos, el sistema identifica una o varias rutas posibles entre origen y destino, descompuestas en arcos del grafo.
	2.1.4 El sistema consulta las velocidades promedio históricas por arco para los arcos que componen las rutas identificadas.
	2.1.5 Si existen datos históricos suficientes, el sistema calcula el tiempo promedio estimado de viaje entre el origen y el destino.
	2.1.6 Si no existen datos históricos suficientes para uno o más arcos de las rutas, el sistema indica la falta de datos y, opcionalmente, presenta una estimación parcial o sugiere una ruta alternativa (si está disponible).
2.2 El usuario selecciona la opción “Estado del sistema de transporte” (y, si aplica, una zona o ruta de interés).	2.2.1 El sistema consulta las métricas históricas de estado del sistema (por ejemplo, velocidades promedio por zona, tiempos promedio por arco relevantes, indicadores de desempeño, número de incidentes históricos).

Juan David Rosero Gómez A00400160  
 Nicolás Jiménez Vargas A00400906  
 Jose Daniel Guzmán Castrillón A00401303

	2.2.2 Si no hay datos históricos suficientes para los filtros seleccionados, el sistema informa que no hay información disponible para esa consulta.
3. El usuario solicita ver los resultados de la consulta.	
	3.1 El sistema prepara la respuesta con la información calculada y/o consultada según el tipo de consulta escogido.
	3.1.1 Para la consulta de tiempo promedio de viaje, el sistema presenta el tiempo promedio estimado de viaje, las rutas consideradas y una explicación básica (por ejemplo, distancia aproximada y número de arcos recorridos).
	3.1.2 Para la consulta de estado del sistema, el sistema presenta un resumen de indicadores históricos (por ejemplo, velocidades promedio por zona, tiempos promedio por arco destacados, niveles de desempeño u otros indicadores configurados).
<b>Poscondición</b>	El usuario visualiza información basada en datos históricos: Un tiempo promedio estimado de viaje entre los puntos de origen y destino, y/o Un resumen del estado del sistema de transporte con indicadores históricos, de acuerdo con los filtros utilizados.
<b>Observaciones</b>	Este caso se apoya en RF6 para los datos almacenados y en RF7/RF19 para los análisis publicados. La interacción puede darse desde una interfaz web La información mostrada corresponde a resultados previamente calculados a partir de datagramas históricos

- **CU01RF6**

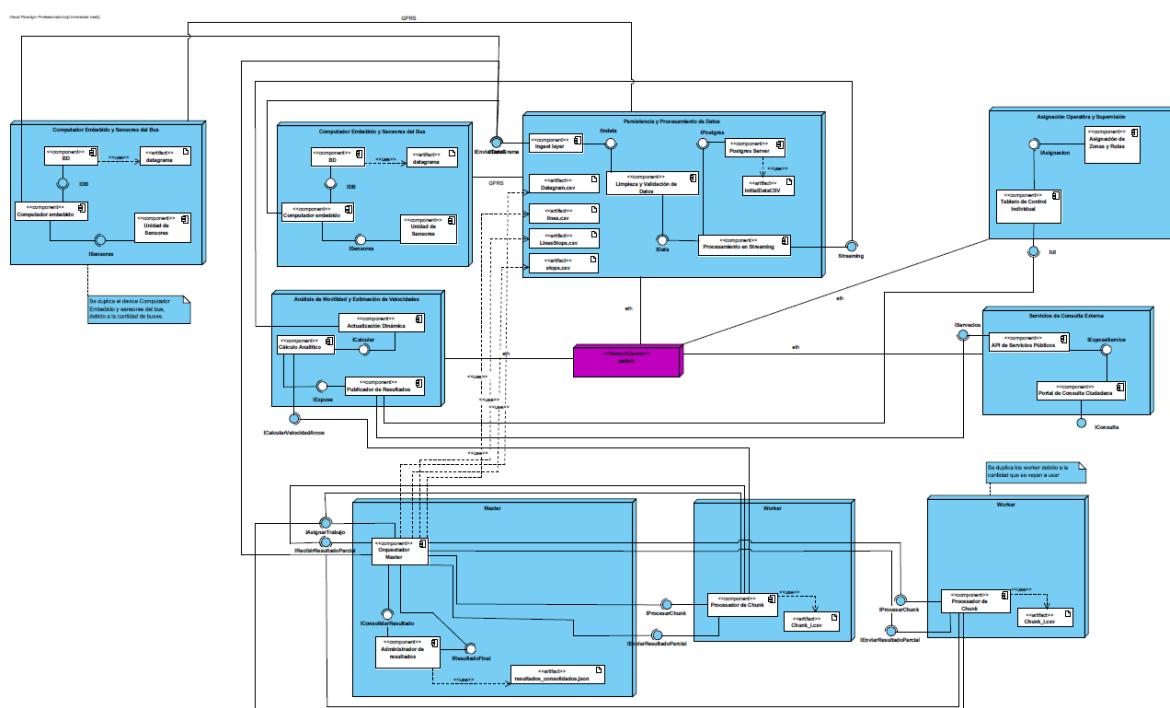
<b>ID</b>	CU01RF06 Cálculo automático de velocidades históricas por arco
<b>Autor</b>	Daniel Guzmán - Juan Rosero - Nicolás Jiménez
<b>Caso de uso</b>	Calcular automáticamente la velocidad promedio histórica de todos los arcos del grafo

Juan David Rosero Gómez A00400160  
 Nicolás Jiménez Vargas A00400906  
 Jose Daniel Guzmán Castrillón A00401303

<b>Casos de uso relacionados</b>	CU01RF09 - CU01RF13
<b>Actor</b>	Sistema y Administrador
<b>Precondición</b>	Los datagramas históricos (eventos de posición GPS y otros eventos relevantes) se encuentran almacenados en el repositorio definido; la base de datos operacional y la base de datos analítica están disponibles; existe una planificación configurada para ejecutar el proceso de análisis de manera automática.
<b>Contexto</b>	El sistema debe procesar de forma automática los datagramas históricos generados por los buses del SITM-MIO para calcular y mantener actualizada la velocidad promedio histórica de todos los arcos del grafo de rutas. Los resultados serán utilizados por otros subsistemas para visualización y consulta.
<b>Acciones del usuario</b>	<b>Acciones del sistema (Incluye flujo normal y excepciones)</b>
1. El proceso programado de análisis se activa según la planificación configurada.	
	2. El sistema inicia la ejecución del proceso automático de análisis histórico.
	3. El sistema identifica los datagramas históricos que aún no han sido procesados o que requieren recalcularse en el repositorio de almacenamiento.
	4. El sistema lee los datagramas por lotes, extrayendo los eventos relevantes (por ejemplo, eventos de posición GPS).
	4.1 Si se detectan datagramas corruptos o inválidos, el sistema registra el error, descarta esos registros y continúa con el resto.
	5. El sistema calcula velocidades instantáneas y tiempos de recorrido por tramo (arco) a partir de las posiciones sucesivas de cada bus.
	6. El sistema agrega los resultados para obtener la velocidad promedio histórica por arco para todos los arcos que cuentan con datos.
	7. El sistema persiste las velocidades promedio por arco y las métricas asociadas en la base de datos analítica, actualizando los valores existentes.

	8. El sistema marca los datagramas procesados como “analizados” para evitar reprocesamiento innecesario.
	9. El sistema muestra el estado del proceso (en ejecución, completado, errores) y un resumen: cantidad de datagramas procesados, número de arcos actualizados y errores encontrados.
1.1 El administrador consulta el estado del proceso de análisis.	
<b>Poscondición</b>	Las velocidades promedio históricas de todos los arcos del grafo quedan calculadas o actualizadas en la base de datos analítica, listas para ser consumidas por los casos de uso de visualización (controladores) y de consulta (usuarios finales).
<b>Observaciones</b>	

## E. Diagrama de deployment



Juan David Rosero Gómez A00400160  
Nicolás Jiménez Vargas A00400906  
Jose Daniel Guzmán Castrillón A00401303