

# Análisis de Movilidad Urbana con Spark

Samuel

Matemáticas Aplicadas a la Computación

FES Acatlán - UNAM

Abril 2025

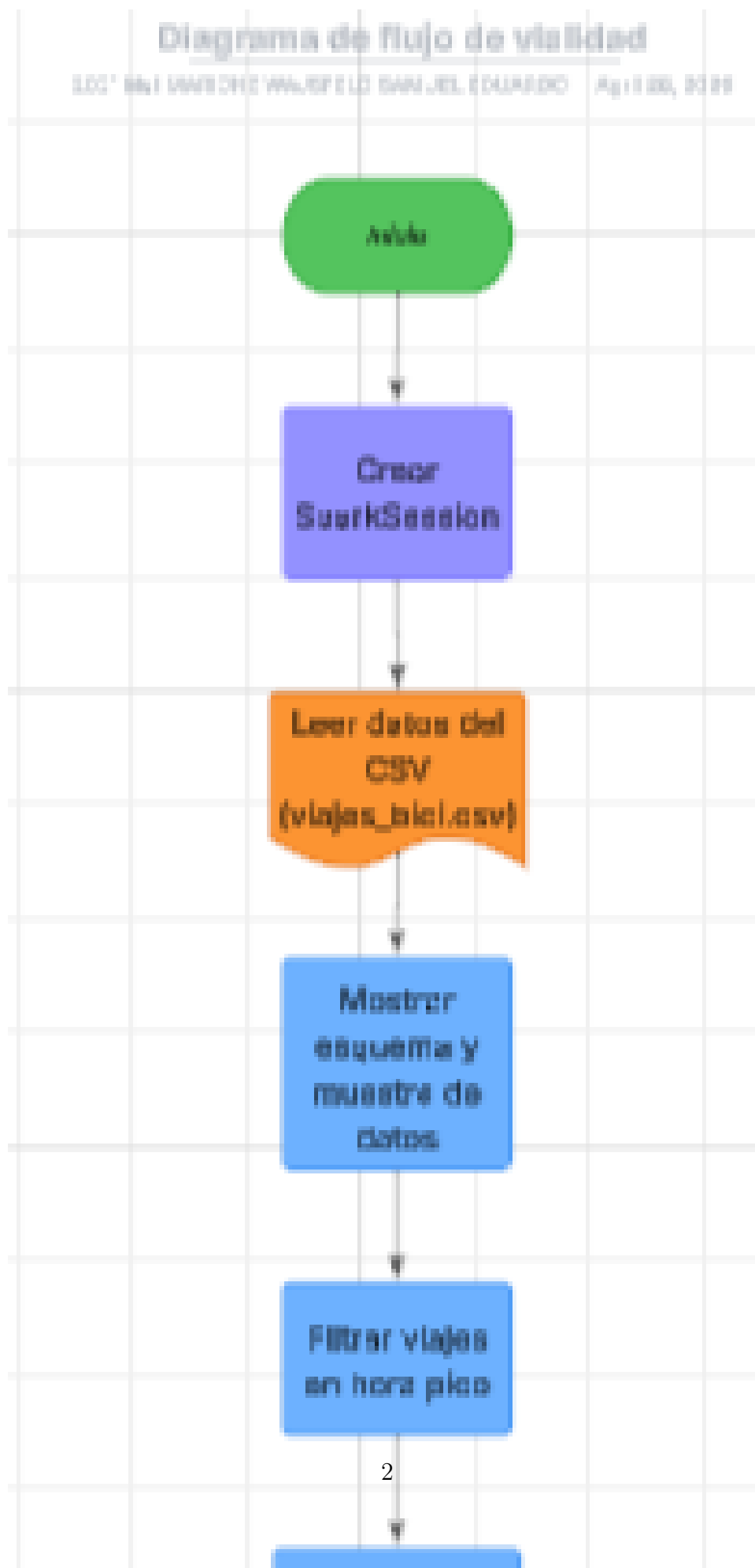
## 1. Introducción

En esta práctica se aplicaron conceptos fundamentales de Spark (RDDs, DataFrames, transformaciones y acciones) para analizar datos de movilidad urbana simulados, simulando registros de viajes de transporte público en la Ciudad de México.

## 2. Dataset

Se utilizó un dataset sintético basado en 10 estaciones principales de la Ciudad de México, generando 1000 viajes en horarios realistas, considerando duración promedio de trayectos entre 5 y 60 minutos.

### 3. Diagrama de flujo



## 4. Operaciones realizadas

- Lectura de datos en CSV.
- Filtro de viajes en hora pico (7–9am, 5–7pm).
- Agrupación de rutas más congestionadas.
- Conteo de viajes por estación de origen y destino.
- Exportación de resultados a CSV.
- Visualización gráfica del resumen.

## 5. Resultados principales

### 5.1 Esquema de datos y muestra

```
[Esquema de datos]
root
|-- viaje_id: integer (nullable = true)
|-- hora_inicio: timestamp (nullable = true)
|-- estacion_origen: string (nullable = true)
|-- estacion_destino: string (nullable = true)
|-- duracion_min: integer (nullable = true)
```

### 5.2 Viajes en hora pico

viaje_id	hora_inicio	estacion_origen	estacion_destino	duracion_min
1	2025-04-28 18:32:00	Tláhuac	Polanco	10
3	2025-04-28 08:24:00	Santa Fe	Tláhuac	46
5	2025-04-28 08:32:00	Polanco	Reforma Norte	55
6	2025-04-28 08:35:00	Coyoacán	Condesa	43
8	2025-04-28 18:36:00	Condesa	Universidad	38
10	2025-04-28 19:09:00	Polanco	Coyoacán	60
11	2025-04-28 19:25:00	Polanco	Tláhuac	9
13	2025-04-28 17:16:00	Condesa	Insurgentes Sur	24
14	2025-04-28 08:02:00	Polanco	Condesa	8
15	2025-04-28 08:54:00	Condesa	Polanco	34
16	2025-04-28 18:19:00	Insurgentes Sur	Chapultepec	31
18	2025-04-28 17:06:00	Santa Fe	Condesa	21
19	2025-04-28 08:52:00	Insurgentes Sur	Universidad	5
20	2025-04-28 19:58:00	Universidad	Tláhuac	52
23	2025-04-28 08:20:00	Centro Histórico	Chapultepec	55
24	2025-04-28 17:25:00	Santa Fe	Tláhuac	58
26	2025-04-28 17:06:00	Reforma Norte	Tláhuac	50
27	2025-04-28 19:50:00	Reforma Norte	Universidad	56
29	2025-04-28 19:45:00	Tláhuac	Chapultepec	37
30	2025-04-28 18:40:00	Condesa	Chapultepec	57

### 5.3 Rutas más populares

[Rutas más populares]		
estacion_origen	estacion_destino	count
Tláhuac	Universidad	17
Coyoacán	Tláhuac	16
Centro Histórico	Tláhuac	15
Tláhuac	Chapultepec	15
Universidad	Reforma Norte	14
Tláhuac	Reforma Norte	14
Polanco	Reforma Norte	14
Condesa	Universidad	13
Santa Fe	Tláhuac	13
Condesa	Polanco	12
Santa Fe	Reforma Norte	12
Centro Histórico	Chapultepec	12
Chapultepec	Coyoacán	12
Santa Fe	Coyoacán	12
Condesa	Coyoacán	12
Santa Fe	Centro Histórico	12
Centro Histórico	Universidad	12
Insurgentes Sur	Centro Histórico	12
Universidad	Coyoacán	12
Insurgentes Sur	Universidad	11

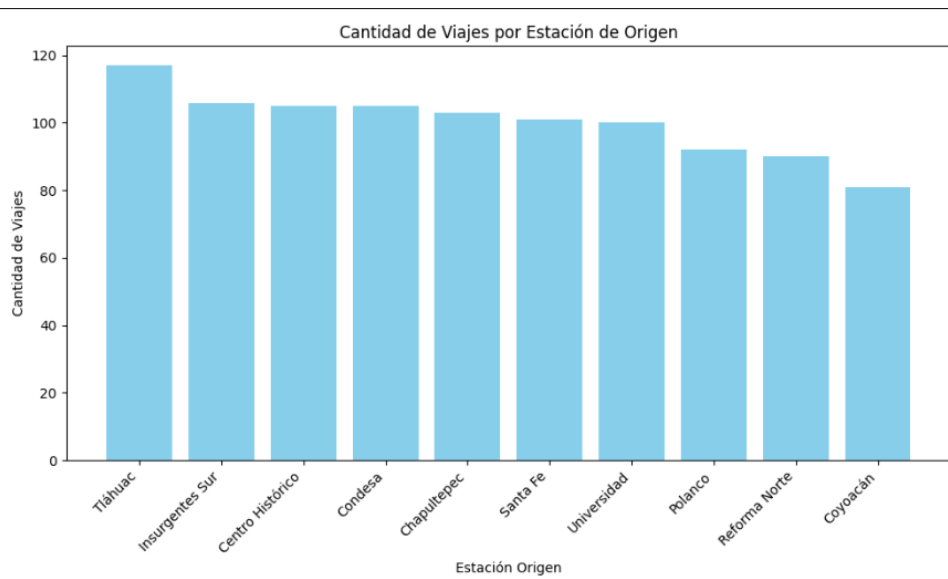
## 5.4 Resumen general

```
[Resumen general de viajes]
```

```
+-----+-----+
| estacion_origen|count|
+-----+-----+
|           Tláhuac|   117|
| Insurgentes Sur|   106|
| Centro Histórico|   105|
|           Condesa|   105|
|    Chapultepec|   103|
|         Santa Fe|   101|
|   Universidad|   100|
|         Polanco|    92|
| Reforma Norte|    90|
|       Coyoacán|    81|
+-----+-----+
```

```
+-----+-----+
| estacion_destino|count|
+-----+-----+
|           Tláhuac|   122|
|       Coyoacán|   114|
| Reforma Norte|   112|
|   Universidad|   111|
| Centro Histórico|    93|
+-----+-----+
```

## 5.5 Gráfica de viajes por origen



## 6. Uso ético de IA

Durante esta práctica se utilizó IA como apoyo para estructurar el análisis, depurar errores de sintaxis y optimizar el flujo de procesamiento, documentando de forma responsable su uso.

## 7. Conclusión

La práctica permitió comprender de forma práctica el procesamiento distribuido de datos reales mediante Spark, evidenciando la facilidad para manipular grandes volúmenes de información de movilidad urbana.