

Conceptos y aplicaciones en Big Data

Trabajo Práctico 3 – Modalidad de cursada presencial

Spark streaming

Pautas generales

- La entrega consiste en la implementación de scripts de Spark, resolviendo todas las consignas presentes en este enunciado.
- Los alumnos pueden conformar grupo de no más de dos integrantes y hacer una única entrega grupal.
- La entrega se realiza por la mensajería del curso en IDEAS.
- La fecha límite de entrega es el 23 de diciembre de 2022.
- La calificación obtenida en este TP será tenida en cuenta en la nota final de la materia.

Enunciado

Una empresa de remises monitorea la geoposición de cada uno de sus vehículos en cada uno de los viajes realizados. La ciudad por donde realizan los viajes los vehículos de la empresa es un cuadrado conformado por 100 avenidas verticales y 100 calles horizontales. Y la geolocalización se transmite del vehículo a la empresa cada vez que el propio vehículo llega a una intersección entre una avenida y una calle.

Los autos generan un stream de datos con su geoposición. El stream posee los siguientes atributos: (ID_vehículo, Avenida, Calle, Timestamp, Destino).

Los scripts en Spark deben permitir el monitoreo de la información pedida en cada enunciado, cada un cierto intervalo de tiempo (por ejemplo, cada 15 segundos). La información solicitada debe ser histórica, desde el comienzo de la escucha del stream.

- 1) Implemente un script de Spark que permita conocer cuántos viajes realizó cada vehículo. Recordar que un viaje es una serie de coordenadas que finalizan en un destino determinado.

Ejemplo:

ID Vehículo	Cantidad de viajes
511	4
697	11
151	1
653	3

- 2) Implemente un script de Spark que permita conocer cuál es la esquina “promedio” entre todas las esquinas cuyo destino es “Otro”.

Ejemplo:

Esquina promedio

(79.25, 25.125)