

Análisis del tráfico de NYC

Se desea implementar una serie de consultas mapreduce que permitan obtener datos estadísticos del tráfico de la ciudad de Nueva York a partir de unos archivos csv que genera el sistema de control de tráfico..

Requisitos de la práctica (Consultas)

Se entregará dos archivos por cada consulta, uno para el mapper y otro para el reducer. Los nombres de los archivos deberán ser mX.py para el mapper y rX.py para el reducer donde X se corresponde con el número de consulta.

El archivo debe poder ser ejecutable desde el terminal, por lo que deberá comenzar con la línea `#!/usr/bin/python`.

El fichero csv contendrá el repositorio de información semi-estructurada sobre la que se realizarán las consultas. Se simulará realizar las consultas sobre cluster a través del archivo. En ningún caso será necesario realizar las consultas sobre Hadoop.

La librería `sys` de Python es necesaria para acceder a la información que se pasa por `stdin` tanto al mapper como al reducer: `sys.stdin`

Se ha de tener en cuenta que es posible que existan errores en los datos proporcionados. En caso de encontrar un error se deberá:

- Más o menos columnas: Se salta la fila.
- Cadenas de texto donde se espera valores numéricos: Se salta la fila.
- Espacios en blanco al principio y al final de una cadena: Cadenas de texto con mismo contenido con o sin espacios antes y/o después de la cadena deben ser equivalentes.

Consultas

1. Tráfico total por calle.
2. Tráfico medio por hora del día.
3. Calle con más tráfico y tráfico total.
4. Hora del día con más tráfico por calle.
5. Calle que más tráfico aporta a Broadway y tráfico total aportado.
6. Calle que más tráfico aporta a Broadway por franja de hora.
7. Diferencia entre el tráfico total de entrada a Broadway frente al de salida.
8. Sentido de cada calle con más tráfico y tráfico total.

Normativa de realización, entrega y evaluación de la práctica:

- La práctica se realizará y entregará por parejas o de forma individual.
- La práctica se realizará en Python..
- La entrega deberá hacerse mediante el campus virtual antes del domingo 28 de mayo de 2017 a las 23:59 horas (hora peninsular en España).
- Las practicas entregadas fuera de plazo, serán calificadas sobre 9. Por cada día de retraso en la entrega se reducirá el rango de calificación en 0,2 puntos.
- La entrega se compondrá de un único fichero ZIP, que contendrá el directorio del proyecto con un listado de las dependencias necesarias.
- Se considerará suspensa toda práctica cuyo fichero comprimido no contenga los ficheros fuente.
- Al hacer entrega del fichero comprimido, éste se habrá renombrado con el nombre y el primer apellido del alumno escritos en letras mayúsculas y separados mediante un guion bajo '_'. Ejemplo: QUIJOTE_DE_LA_MANCHA.zip
- Cualquier sospecha de COPIA entre dos o más prácticas o de código obtenido en internet derivará en la calificación de 0 para todos los alumnos involucrados.
-