

**Tipo** : Enunciado de laboratorio  
**Capítulo** : Componentes de Hadoop  
**Duración** : 30 minutos

---

## **I. OBJETIVO**

Conocer cómo implementar los componentes de Hadoop.

## **II. REQUISITOS**

Los siguientes elementos de software son necesarios para la realización del laboratorio:

- Virtual Box
- Hortonworks Sandbox for HDP
- Putty
- Winscp

## **III. EJECUCIÓN DEL LABORATORIO**

1. Ingresar al servidor Hortonworks,  
Para esto deberá ingresar a la siguiente dirección:  
<http://127.0.0.1:8080>
2. Cargar los archivos de ejemplo.
3. Crear el script en Pig.
4. Realizar la carga de datos del archivo.
5. Establecer un filtro de datos
6. Establecer un filtro de datos
7. Realizar operaciones con la data del archivo timesheet
8. Filtrar los datos por cada DriverId
9. Sumar las horas y millas por cada Driver
10. Realizar el JOIN de los campos DRIVERID, NAME, HOURS AND MILES LOGGED
11. Ejecutar y grabar el script
12. Código final del script
13. Ejecutar el script utilizando TEZ.