

**Tipo** : Enunciado de laboratorio  
**Capítulo** : Spark  
**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. Antes de iniciar este laboratorio deberá haber realizado el laboratorio 6.6 de este curso.
2. Ingresar al servidor Hortonworks, para esto deberá ingresar a la siguiente dirección:  
<http://127.0.0.1:8080>
3. Abra Ambari y asegúrese de que **Zeppelin Notebook** esté funcionando (Pantalla de la izquierda). Como se muestra en la captura de pantalla a continuación, use el menú desplegable "Quick Links" para acceder a la interfaz de usuario de Zeppelin.
4. Una vez que la interfaz de usuario de Zeppelin esté abierta, haga clic en "Import Note".
5. Importar ClickstreamAnalytics.json, una vez que se abre Zeppelin, haga clic en el ícono correcto en la barra de navegación para mostrar el código que acompaña a los datos visualizados. Vea la siguiente captura de pantalla para la ubicación de este ícono.
6. Identificar el estado con la mayoría de los clientes. Echemos un vistazo al primer gráfico en el Notebook.
7. Entender la demografía del cliente. Desplácese hacia abajo y revise la siguiente sección con un gráfico. Profundicemos un poco más y veamos cómo logramos la visualización. Escriba la consulta para filtrar los datos demográficos (age, gender, category)
8. Analizar la distribución de la categoría de interés. Finalmente, veamos el último gráfico en este cuaderno. Parece que la ropa (clouting) es claramente la razón más popular por la que los clientes visitan el sitio web.