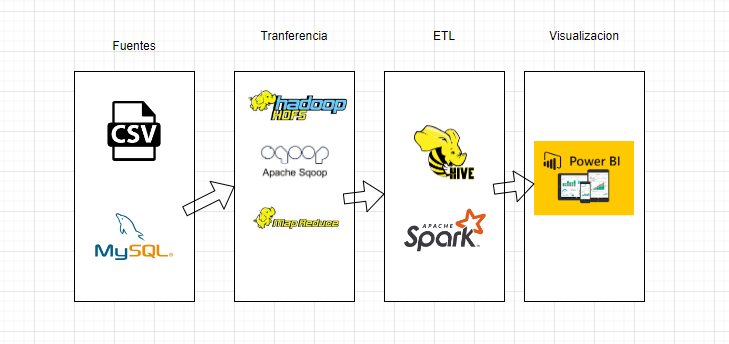
**PROYECTO FINAL PARA LA RUTA DE APRENDIZAJE DE DATA ENGINEER  
  
No mecanizarnos:**Antes de realizar el proyecto y tener claro los procedimientos y pasos a realizar primero pongámonos a pensar realmente el porque de realizar todo lo que hacemos, cual es el propósito y trasfondo que existe detrás de toda esta solución, la importancia de disponibilizar la información a los usuarios finales es esencial para que ellos puedan realizar un análisis y puedan tomar decisiones acertadas.   
  
Todo esto es posible si los Data Engineer realizan su trabajo de forma excepcional, ya que a medida que el volumen de la data crece, las fuentes de orígenes de información aumentan y se vuelve ingobernable y difícil de unificar. Por esto hoy en día existen diferentes herramientas como Hadoop(On premisse) o Cloud las cuales nos permiten agilizar todo este proceso de Big data y llevar a los usuarios finales la información que necesiten.  
  
Basándonos en haber recorrido también la ruta de aprendizaje de Data Analyst, con este proyecto no solo buscaremos una solución en hadoop sino también el valor que se aporta al negocio, se solicitará tener la data que hayamos podido trabajar en hadoop en una herramienta de análisis y visualización como Power BI u alguna otra y mostrar un algún indicador de valor para el negocio. Recordarles que están trabajando en equipos y pueden apoyarse unos a otros durante todo el proceso.  
  
**PAUTAS Y RECOMENDACIONES:**  
Para este proyecto pueden utilizar todo lo aprendido hasta el momento o incluso agregar los conocimientos que posean para complementarlo.  
  
Siéntanse en completa libertar de aplicar todas las herramientas que crean conveniente y necesarias en cualquier fase del proyecto, en donde sientan que puedan aplicar lo aprendido hasta el momento (Al final se dará un feedback si realmente fue una manera óptima de solución), ustedes cuentan con esa libertad ( SSIS , ETL, Python, Spark, Power BI, Power Query, Hive, Sqoop, chatGPT, etc)  
  
**CRITERIOS A EVALUAR PROYECTO FINAL:**  
Se presentará un zip con todos los adjuntables posibles incluyendo un Word con las siguientes criterios.  
1) Integrantes del grupo, solo los que hayan formado parte del proyecto  
2) Alcance: “Que se busca y hasta donde se piensa llegar” (DETALLAR CUAL ES LA SOLUCIÓN)  
3) Diagrama de la solución: Pueden usar cualquier herramienta como Paint , Drawio, PPT para hacer un diagrama General de la solución.  
Ejemplo :



4) Evidenciar haber utilizado el ecosistema de hadoop

5) Algunas capturas de los pasos realizados ( comandos digitados , tranferencias de archivos, la data en hdfs, las tablas en hive, la conexión con BI, etc)

5) Conclusiones del proyecto.

**PROYECTO:**

Para este proyecto se necesita llevar data que poseamos de alguna organización ficticia hacia el ecosistema de **hadoop, en donde tenemos que realizar un proceso de migración** y si desean conveniente realizar el proceso ETL dentro de hadoop queda en libertad de cada uno. **Posteriormente se debe contar con la data en alguna herramienta que nos pueda indicar un análisis,** para esto pueden apoyarse en una herramienta BI o exportar la información para **luego mostrar un indicador de valor que ayude al negocio**.  
  
Cumpliendo con lo mencionado ya tendrían el proyecto final listo y contar con su certificado de Data Engineer con Hadoop.  
  
Para realizar este proyecto tienen 2 opciones:

**Opción1**

Trabajar con el dataset , bd , txt , etc que deseen conveniente y cumplir con el objetivo planteado. Recuerden que todo lo define su alcance y tiene que tener sustento del porque se realiza la solución.

**Opción2**

En caso de no contar con un dataset propio realizar las siguientes pautas:  
  
En la maquina virtual ya contamos con un gesto de datos mysql en el cual vamos a importar una bd mysql muy conocida llamada “sakila” la cual es una base de datos de una empresa que se dedica al alquiler de películas.   
  
Pasos a tomar en consideración :  
  
1) Se tendrán los archivos “sakila-schema.sql” y “sakila-data.sql” para importar la base de datos mencionada.  
  
2) Se tendrán que llevar estos archivos a home/maría\_dev   
  
3) Se tendrá que ingresar al mysql de la maquina virtual de la forma en la que se mencionó en las sesiones.  
4) Se creará la base de datos ejecutando la siguiente sentencia “créate database sakila;”

5) Se ejecutará el siguiente comando dentro de mysql para crear las tablas “SOURCE /home/maria\_dev/sakila-schema.sql;”

6) Se ejecutará el siguiente comando dentro de mysql para poblar las tablas “SOURCE /home/maria\_dev/sakila-data.sql;”  
  
7) Analizar la base de datos(investigando en internet) y migrar más de 4 tablas a HIVE ( Tienen que ser la tablas que consideren necesarias para posteriormente ser utilizadas para realizar un análisis o un ETL lo que usted crea conveniente)  
  
8) Lo ideal es que realice esta migración con Sqoop pero adicionalmente les estoy compartiendo la base de datos en formato CSV y puedan apoyarse en caso desean migrar alguna las tablas con comandos hdfs. (Recuerden especificar el delimitador de coma a la hora de crear la tabla en hive si en caso deciden optar por este medio)

9 ) Recordar y tener en consideración los tipos de datos semejantes con los que se deben de crear las variables en hive (string , int , etc ) para que cuando realicen la transferencia no tengan inconvenientes.

10) Llevar la información a una herramienta de visualización y realizar un análisis de valor con la información que se ha trabajado. (Dashboard , kpi, etc)

Antes que nada, agradecerles por tomarse la dedicación y el interés de leer este proyecto, siéntanse en total libertad de realizarlo como mejor lo crean conveniente, en el classroom encontrarán todos los videos y materiales necesarios para realizarlo.

Si tuvieran alguna duda o consulta pueden escribirme con toda confianza a mi número es el 933398057 o mi linkedin <https://www.linkedin.com/in/rodrigojauregui/>   
  
  
Si en caso no llegasen a enviar el proyecto en el tiempo establecido y desean realizarlo más adelante igual también pueden consultarme, la idea es que podamos aprender entre todos y apoyarnos.  
  
  
Saludos,  
  
Rodrigo Jáuregui