# PROYECTO 1

Hemos visto que Hadoop es un ecosistema que consta de diferentes herramientas para almacenar, procesar y analizar datos Big Data (Petabytes: 1015 byte).

Hasta ahora hemos aprendido cómo instalar un clúster Hadoop en máquinas virtuales, así como en AWS y cómo lanzar tareas utilizando el motor MapReduce.

Para avanzar en el estudio de las herramientas del ecosistema Hadoop en este proyecto nos centraremos en Hive (SQL), HBase(NoSQL) y HUE (interfaz gráfica).

Lo que vamos a hacer es descargarnos desde Kaggle algunos ficheros, subirlos a Hadoop y utilizando Hive y HBase haremos consultas sobre sus datos, la herramienta HUE la podemos usar como interfaz.

Los pasos a seguir son:

1.- Crear un clúster en AWS con las siguientes aplicaciones: Hadoop, HBase, Hive y Hue.

2.- Descargar un repositorio de Kaggle.

3.- Subir los archivos Hadoop (AWS)

4.- Crear un pequeño manual de las instancias CRUD HBase y Hive.

5.- Utilizando Hive y HBase realizar consultas acerca de estos archivos subidos.

5.1.- Estudiar los datos almacenados en los archivos para crear las tablas en las que cargar los datos.

5.2.- Crear las tablas en HBase y Hive.

5.3.- Cargar los datos de los ficheros en las tablas.

5.4.- Crear diferentes consultas sobre las tablas y mostrar el resultado obtenido.

Todo tiene que estar documentado con ejemplos ejecutados en vuestro propio equipo o máquina virtual:

* Código a ejecutar en formato texto (para poder cortar y pegar)
* Pantallazo del código ejecutado

## CRITERIOS DE EVALUACIÓN

* (8p) Documentación:
  + Portada (0.4p)
  + Índice con número de página (enlaces) (0.4p)
  + Número de página(0.4p)
  + Buena presentación (alineaciones, espacios en blanco innecesarios, imágenes no colocadas correctamente) (0.4p)
  + Faltas de ortografía (0.4p)
  + En cuanto a su contenido se valorará: (6p)

Si contiene todos los puntos. (4p)

El lenguaje utilizado: Si se utilizan los términos técnicos adecuadamente.(2p)