

Sigla Asignatura	BDY7101	Nombre Asignatura	Big Data	Tiempo	2h
Experiencia de Aprendizaje N° 1	Aprende qué es Big Data				
Actividad N° 1.2	Actividad de uso de Docker				
Nombre del Recurso Didáctico	1.2.2 Actividad de uso de Docker				

1. APRENDIZAJES E INDICADORES DE LOGRO

Aprendizajes (Procedimentales, Actitudinales y conceptuales)	Indicadores de logro
Seleccionar el tipo de infraestructura de Big Data, según características y ubicación de la organización y/o problema.	Realiza una propuesta de uso de infraestructura de solución de Big Data para la organización.
Es riguroso con la aplicación de las herramientas para lograr el mayor desempeño de la misma.	Es capaz de analizar los resultados obtenidos de la aplicación de las herramientas del ecosistema Hadoop y/o Spark para poder formular conclusiones que resuelvan el problema.
Describe los aspectos de la infraestructura de Big Data, de acuerdo a características y necesidades de la organización.	Reconoce las funciones y características del entorno de Big Data, de acuerdo a características y necesidades de la organización.

2. DESCRIPCIÓN GENERAL ACTIVIDAD

Esta actividad tiene **carácter formativo**, cuya finalidad es visualizar lo que aprendes, en la directa medida que tu docente de asignatura te va retroalimentando constantemente, tanto a nivel individual como colectivo (equipo de trabajo).

Contexto

Apache Hive es una infraestructura de almacenamiento de datos construida sobre Hadoop, la cual soporta el análisis de grandes conjuntos de datos almacenados bajo HDFS de Hadoop.

Apache Hive inicialmente fue desarrollado por Facebook y, actualmente utilizado por grandes compañías como los son Netflix y Financial Industry Regulatory Authority (FINRA). Además, también es necesario considerar que Amazon mantiene una derivación de Apache Hive dentro de sus servicios Amazon Web Services.

En tanto, *Apache Hive* ofrece un lenguaje de consultas basado en *SQL* llamado *HiveQL* con esquemas para leer y ejecutar consultas en *MapReduce, Apache y Apache Spark*. Por último, debes considerar que estos últimos tres motores de ejecución corren bajo *YARN*.

Paso 1

Docente Diseñador	Gerardo Cerda Neumann	Revisor metodológico	Manuela Jiménez
-------------------	-----------------------	----------------------	-----------------



Utilice *Dockers* para instalar un contenedor de *Hive*, la recomendación es utilizar la imagen *nagasuga/docker-hive* Utiliza Docker Hub (http://hub.docker.com) para buscar esta imagen.

Paso 2

Una vez realizado el paso anterior deberá realizar las siguientes tareas:

- 1. Crear la base de datos EJEMPLO
- 2. Dentro de la base de datos EJJEMPLO, crear la tabla PERSONA, la cual tendrá dos atributos. A saber: Nombre como string y Rut como entero.
- 3. Realizado los pasos anteriores, ejecute el comando DESCRIBE para ver los atributos de la tabla PERSONA.
- 4. Finalmente, ejecute el comando DESCRIBE FORMATTED con tal de visualizar las propiedades de la tabla PERSONA

Paso 3

Habiendo realizado los pasos anteriores, responda los siguiente: ¿Qué características de DOCKER permiten la creación de una infraestructura de pruebas para proyectos Big Data?

Aspectos a considerar en la evaluación

- I. Identifica todos los elementos de Docker en la situación analizada
- II. Identifica todas las características útiles de Docker
- III. Identifica características útiles de Docker para implementar una solución de Big Data
- IV. Propone argumentos convincentes para la utilización de Docker.
- V. Identifica la necesidad de resolver la problemática mediante la infraestructura que proporciona Docker.
- VI. Concluye respecto al beneficio de usar Docker como parte de la solución
- VII. Analiza el aporte de Docker para construir una solución de Big Data.
- VIII. Identifica situaciones factibles de ser resueltas mediante Big Data que Dockers permita solucionar. Debe argumentar la utilidad de Docker en cada contexto.