

/

<b>Nombre:</b> Jesús Alejandro Valencia Valadez		<b>Matrícula:</b> 3013480
<b>Nombre del curso:</b> Fundamentos para programación de Big Data	<b>Nombre del profesor:</b> Sergio Arturo Damián Sandoval	
<b>Módulo:</b> 1	<b>Actividad:</b> Actividad 2	
<b>Fecha:</b> 05/02/22		

1. Investiga y describe qué son y para qué se utilizan cada una de las siguientes tecnologías:

A. SQL

Es un lenguaje de programación para almacenar y procesar información de una base de datos esta base puede utilizar SQL para almacenar, actualizar, eliminar, buscar y recuperar información, también es utilizada para mantener optimizar el rendimiento de la base de datos.

B. Python

Lenguaje de programación usado principalmente para programación orientada a objetos sin embargo tiene miles de módulos de terceros por lo que le permite ser un lenguaje con infinitas posibilidades como desarrollo web, base de datos, numérico, Juegos, programación de red entre otras.

C. R

Es un entorno y lenguaje de programación abierto, libre y gratis que proporciona variedad de técnicas estadísticas y gráficas, al ser un software libre al igual que Python le permite tener infinitos usos como este mencionado.

D. Anaconda

Es una distribución de lenguajes de programación Python y R para computación científica la cual tiene como ventaja simplificar la gestión e implementación de sus paquetes para todas las distribuciones de SO.

#### E. Google Colaboratory

Permite a cualquier usuario escribir y ejecutar código de Python en el navegador de manera gratuita, es ideal para aprendizaje, análisis de datos y educación.

#### F. Pandas

Es una biblioteca para Python la cual ofrece estructuras de datos y operaciones para manipular tablas numéricas y series temporales la cual resulta gracias a estos atributos ser una herramienta buena para Data Science.

#### G. Hadoop

Es un framework que permite el procesamiento distribuido de grandes conjuntos de datos utilizando modelos de programación simples, el cual está hecho para poder trabajar desde servidores individuales hasta miles de máquinas.

#### H. Apache Spark

Es un multi lenguaje para ingeniería de datos, ciencia de datos y aprendizaje automático en máquinas, algunas características son el análisis de datos SQL, el procesamiento de lotes de datos, ciencia de datos a escala.

Contesta las siguientes preguntas:

- I. ¿En qué casos es más recomendable utilizar Anaconda y en qué casos es más recomendable usar Google Colaboratory?

Anaconda la utilizaría para trabajar en un entorno Python local y controlar accesos a paquetes y versiones específicas, también si se requiere la capacidad de procesamiento y almacenamiento mas grande de la que puede ofrecer Google Colaboratory.

Google Colaboratory lo utilizaría cuando se requiere accesos rápido y fácil a un entorno de desarrollo de Python sin tener que instalar o configurar nada de manera local, también para el trabajo de proyectos chicos o de aprendizaje para compartir con demás personas.

J. ¿En qué casos preferirías utilizar Python en vez de R y viceversa?

Python lo preferiría en caso de integración con otros sistemas y aplicaciones y teniendo en cuenta que se requiere una mayor experiencia programando que con R

R lo preferiría en caso de un enfoque mas estadístico y grafico en caso de analizar datos y en general cuando tenga que trabajar con estadística gracias a su gama de paquetes.

K. ¿Qué otros lenguajes de programación son normalmente utilizados para la ciencia de datos y el Big Data?

Scala es un lenguaje el cual se ejecuta en la plataforma de Apache Spark.

Juolia diseñado para aplicaciones científicas y análisis de datos

SAS

MATLAB en un entorno de análisis de datos y cálculo numérico.

Fuentes:

*¿Qué es SQL? - SQL - AWS.* (2023). Amazon Web Services, Inc.

[https://aws.amazon.com/es/what-is/sql/#:~:text=El%20lenguaje%20de%20consulta%20estructurada%20\(SQL\)%20es%20un%20lenguaje%20est%C3%A1ndar,relacional%20que%20utiliza%20consultas%20SQL](https://aws.amazon.com/es/what-is/sql/#:~:text=El%20lenguaje%20de%20consulta%20estructurada%20(SQL)%20es%20un%20lenguaje%20est%C3%A1ndar,relacional%20que%20utiliza%20consultas%20SQL).

*Welcome to Python.org.* (2023, January 26). Python.org; Python.org.

<https://www.python.org/about/>

Datademia. (2021, October 18). *¿Qué es R? ¿Para qué sirve? - Datademia.*

Datademia. <https://datademia.es/blog/que-es-r>

Izary Rondón. (2022, February 4). *¿Qué es Anaconda? -Escuela Internacional de*

*Posgrados. Másteres Online N° 1 Empleabilidad.* <https://eiposgrados.com/blog-python/que-es-anaconda/>

*Google Colab.* (2014). Google.com.

<https://research.google.com/colaboratory/intl/es/faq.html#:~:text=Colaboratory%2C%20o%20%22Colab%22%20para,an%C3%A1lisis%20de%20datos%20y%20educaci%C3%B3n>.

*Pandas: herramienta básica para el Data Science.* (2018, June). Máster En Data

Science. <https://www.master-data-scientist.com/pandas-herramienta-data-science/#:~:text=%C2%BFQu%C3%A9%20es%20pandas%3F,tablas%20num%C3%A9ricas%20y%20series%20temporales>.

*Apache Hadoop.* (2022). Apache.org. <https://hadoop.apache.org/>

*Apache Spark™ - Unified Engine for large-scale data analytics.* (2018). Apache.org.

<https://spark.apache.org/>