	<b>ESCUELA DE INGENIERÍA DE SISTEMAS Y COMPUTACIÓN</b>		
<b>Curso:</b>	Minería de Datos y Big Data	<b>Grupo:</b>	Único
<b>Docente:</b>	Mgtr. Ing. Mariana Chavarry Chankay		
<b>Semestre:</b>	2023-2	<b>Fecha de Evaluación:</b>	4 / 12 / 2023
<b>Tipo de Evaluación:</b>	Examen de unidad 3	<b>Duración:</b>	60 minutos

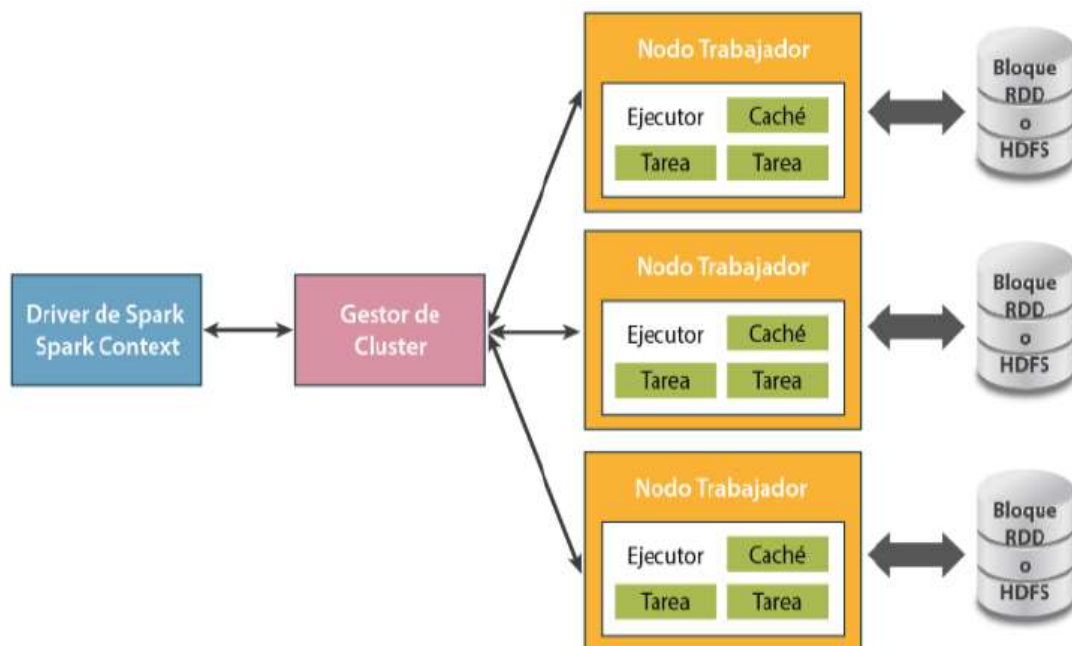
## 1. Instrucciones

### Importante:

- La evidencia de plagio anula la evaluación.
- Enviar el desarrollo de la evaluación dentro de una carpeta comprimida a través del aula virtual.

### Parte I: (8ptos)

La imagen muestra una representación de la arquitectura de **Apache Spark**, en base a ello elabora el algoritmo que permita explicar de acuerdo a cada uno de sus componentes el funcionamiento de Apache Spark en cualquier escenario Big Data presentado.



### Parte II: (12ptos)

Desarrollar lo solicitado aplicando **Spark SQL en Python** haciendo uso de los Datasets que se encuentran en el archivo comprimido **Base\_datos.zip**:

1. Generar una consulta que muestre varias columnas con condiciones aplicando agrupaciones (group by) combinadas con medidas resumen, además se deben de mostrar los registros ordenados según alguna columna.  
Se debe de especificar el enunciado de la consulta desarrollada, la cual deberá de ser de gran utilidad. (4ptos.)

"Siempre parece imposible hasta que se hace"

2. Generar una consulta que muestre varias columnas donde se haga uso de por lo menos 2 datasets, se deberán de mostrar todos los registros del dataset de la derecha estén o no relacionados con el dataset de la izquierda.  
Se debe de especificar el enunciado de la consulta desarrollada, la cual deberá de ser de gran utilidad. **(4ptos.)**
3. Generar una consulta que muestre varias columnas donde se haga uso de por lo menos 2 datasets, que cumplan con al menos 3 condiciones, agrupados (group by) por criterios según medidas resumen.  
Se debe de especificar el enunciado de la consulta desarrollada, la cual deberá de ser de gran utilidad. **(4ptos.)**