

PRIMER APELLIDO	SEGUNDO APELLIDO	NOMBRE DEL ALUMNO/A	DÍA NACIMIENTO	MES NACIMIENTO
M	D	J	04	11

ESTUDIOS	EVALUACIÓN / CONVOCATORIA	AÑO ACADÉMICO
Curso de Especialización en Inteligencia Artificial y Big Data	Final Extraordinaria	
MÓDULO/MATERIA		2024-25
Big Data Aplicado		

### Instrucciones para rellenar el encabezado con tus datos:

- En cada casilla, escribe la primera letra de tu primer apellido, la de tu segundo apellido y la de tu nombre.
- Si tienes un nombre o apellido compuesto, incluye solo la primera letra. Por ejemplo, para:  
*Pérez García, María Isabel*, el código sería: **PGM**  
*García De la Hoz, Rebeca*, el código sería: **GDR**  
*De la Hoz García, María*, el código sería: **DGM**
- En el mes de nacimiento, escribe el mes de nacimiento con número. Por ejemplo, **enero:1**, **diciembre:12**.

### INSTRUCCIONES PARA LA REALIZACIÓN DE LA PRUEBA:

- Este examen dura 90 minutos y debe realizarlo sin ningún tipo de ayuda.
- Debe ser contestado con bolígrafo azul o negro (no lápiz ni ningún otro color).
- El examen consta de 20 preguntas de tipo test.
- Cada pregunta tiene cuatro respuestas posibles y solo una correcta.
- Cada pregunta respondida correctamente se valorará con +0,5 puntos.
- Cada pregunta respondida incorrectamente se valorará con -0,2 puntos.
- Las preguntas no contestadas ni sumarán ni restarán puntos.
- Las contestaciones deben hacerse en la tabla que figura en la siguiente página.
- **NO SE CORREGIRÁN CONSTESTACIONES QUE NO ESTÉN EN LA TABLA.**
- Para contestar, marca la letra de la respuesta que consideres correcta en el recuadro adecuado.
- Se quieres cambiar una respuesta, tacha la respuesta anterior y marca la nueva letra de forma que no haya ambigüedad en la respuesta elegida.
- Si no está clara la respuesta consignada se interpretará que la pregunta no ha sido contestada.

PRIMER APELLIDO <b>M</b>	SEGUNDO APELLIDO <b>D</b>	NOMBRE DEL ALUMNO/A <b>J</b>	DÍA NACIMIENTO <b>04</b>	MES NACIMIENTO <b>11</b>
-----------------------------	------------------------------	---------------------------------	-----------------------------	-----------------------------

### TABLA DE RESPUESTAS

PREGUNTA 1	A	B	<b>C</b>	D	PREGUNTA 11	A	B	C	<b>D</b>
PREGUNTA 2	A	<b>B</b>	C	D	PREGUNTA 12	A	<b>B</b>	C	D
PREGUNTA 3	<b>A</b>	B	C	D	PREGUNTA 13	<b>A</b>	B	C	D
PREGUNTA 4	A	<b>B</b>	C	D	PREGUNTA 14	A	B	<b>C</b>	D
PREGUNTA 5	A	B	<b>C</b>	D	PREGUNTA 15	A	B	<b>C</b>	D
PREGUNTA 6	<b>A</b>	B	C	D	PREGUNTA 16	A	B	C	<b>D</b>
PREGUNTA 7	A	B	<b>C</b>	D	PREGUNTA 17	A	B	<b>C</b>	D
PREGUNTA 8	A	B	<b>C</b>	D	PREGUNTA 18	A	<b>B</b>	C	D
PREGUNTA 9	A	B	<b>C</b>	D	PREGUNTA 19	<b>A</b>	B	C	D
PREGUNTA 10	<b>A</b>	B	C	D	PREGUNTA 20	A	<b>B</b>	C	D

PRIMER APELLIDO	SEGUNDO APELLIDO	NOMBRE DEL ALUMNO/A	DÍA NACIMIENTO	MES NACIMIENTO
M	D	J	04	11

1.- ¿Cómo se llama el nodo que coordina la ejecución de trabajos en YARN?:

- A) NameNode.
- B) Scheduler.
- C) ResourceManager.
- D) NodeManager.

2.- ¿Cuál es la función del paso de "combine" en MapReduce?

- A) Realizar una operación de mapeo en los datos de entrada.
- B) Reducir los datos de salida del "mapper" antes de enviarlos al "reducer".
- C) Reducir los datos de salida del "reducer" antes de enviarlos al "mapper".
- D) Reducir los datos de salida del "reducer" antes de almacenar el resultado en disco.

3.- ¿Cuál es la función del paso de "shuffle" en MapReduce?

- A) Ordenar los datos de salida de los nodos "map" antes de enviarlos a los nodos "reduce".
- B) Dividir los datos de entrada en trozos más pequeños para procesarlos de forma distribuida.
- C) Hacer un muestreo aleatorio simple de los datos.
- D) Realizar una operación de mapeo en los datos de entrada.

4.- ¿Qué componente de Hadoop permite sincronizar el estado de los diferentes servicios distribuidos de Hadoop?

- A) Apache Hive.
- B) Apache ZooKeeper.
- C) Apache HBase.
- D) Apache Flume.

5.- ¿Qué es el factor de replicación en HDFS?

- A) El número de nodos que forman el clúster de Hadoop.
- B) El número de bloques que contiene un fichero en HDFS.
- C) El número de réplicas de cada bloque en diferentes nodos.
- D) El número de ficheros almacenados en un directorio en HDFS.

6.- ¿Qué es lo que realiza el cliente cuando desea leer un fichero de HDFS?

- A) Realiza una llamada al Namenode para obtener los bloques del fichero.
- B) Realiza una llamada al Datanode para obtener los bloques del fichero.
- C) Realiza una llamada al cliente para obtener los bloques del fichero.
- D) Realiza una llamada al servidor para obtener los bloques del fichero.

7.- ¿Cómo se llama el sistema de almacenamiento de Hadoop?:

- A) GFS.
- B) DBFS.
- C) HDFS.
- D) FS.



PRIMER APELLIDO <b>M</b>	SEGUNDO APELLIDO <b>D</b>	NOMBRE DEL ALUMNO/A <b>J</b>	DÍA NACIMIENTO <b>04</b>	MES NACIMIENTO <b>11</b>
-----------------------------	------------------------------	---------------------------------	-----------------------------	-----------------------------

8.- ¿En qué nodo se ejecuta el ApplicationMaster?

- A) En el nodo maestro.
- B) En el Datanode.
- C) En un nodo worker.
- D) En el ResourceManager.

9.- ¿Qué comando HDFS se utiliza para copiar un archivo local al sistema de archivos distribuido de Hadoop?

- A) `hadoop fs -ls`
- B) `hadoop fs -mkdir`
- C) `hadoop fs -put`
- D) `hadoop fs -get`

10.- En YARN, ¿cuál es el equivalente del Namenode en HDFS?

- A) ResourceManager
- B) Datanode
- C) NodeManager
- D) DataManager

11.- ¿Qué operador se utiliza para proyectar o seleccionar columnas de datos en Apache Pig?

- A) SELECT
- B) JOIN
- C) COGROUP
- D) FOREACH

12.- Señala si son o no ciertas las siguientes afirmaciones:

I.- Las consultas de Pig se traducen a procesos MapReduce.

II.- Los datos en Pig se almacenan en una estructura relacional.

- A) I cierta, II cierta
- B) I cierta, II falsa
- C) I falsa, II cierta
- D) I falsa, II falsa

13.- ¿Cuál es el método que se utiliza para emitir pares clave-valor en el mapper de MRJob?

- A) `yield (key, value)`
- B) `output.collect(key, value)`
- C) `emit(key, value)`
- D) `print(key, value)`

PRIMER APELLIDO	SEGUNDO APELLIDO	NOMBRE DEL ALUMNO/A	DÍA NACIMIENTO	MES NACIMIENTO
M	D	J	04	11

14.- Señala si son o no ciertas las siguientes afirmaciones:

- I.- En "dataframes" de Spark las acciones son perezosas.  
II.- En "dataframes" de Spark las transformaciones son perezosas.

- A) I cierta, II cierta  
B) I cierta, II falsa  
C) I falsa, II cierta  
D) I falsa, II falsa

15.-Cuál es el lenguaje nativo con el que está construido Apache Spark:

- A) Python  
B) Java  
C) Scala  
D) SQL

16.- Señala si son o no ciertas las siguientes afirmaciones:

- I.- En MRJob es obligatorio definir un reducer.  
II.- En MRJob puede haber más de un reducer.

- A) I cierta, II cierta  
B) I cierta, II falsa  
C) I falsa, II cierta  
D) I falsa, II falsa

17.- Dado el siguiente fichero "notas.txt", que almacena información no estructurada, indica si son o no ciertas las siguientes afirmaciones:

```
%%writefile notas.txt
pedro 6 7
luis 0 4
ana 7
pedro 8 1 3
```

- I.- Se puede procesar directamente con Hive.  
II.- Se puede procesar directamente con Spark.

- A) I cierta, II cierta  
B) I cierta, II falsa  
C) I falsa, II cierta  
D) I falsa, II falsa

PRIMER APELLIDO <b>M</b>	SEGUNDO APELLIDO <b>D</b>	NOMBRE DEL ALUMNO/A <b>J</b>	DÍA NACIMIENTO <b>04</b>	MES NACIMIENTO <b>11</b>
-----------------------------	------------------------------	---------------------------------	-----------------------------	-----------------------------

18.- Señala si son o no ciertas las siguientes afirmaciones sobre la siguiente consulta:

`SELECT * FROM airports LIMIT 10`

- I.- La sintaxis es válida en Spark.
- II.- La sintaxis es válida en Pig.

- A) I cierta, II cierta
- B) I cierta, II falsa
- C) I falsa, II cierta
- D) I falsa, II falsa

19.- Señala si son o no ciertas las siguientes afirmaciones:

- I.- En un Data Lake, el esquema de los datos se fija en la lectura, no en la escritura.
- II.- Un Datawarehouse requiere conocer bien la estructura de los datos antes de su escritura.

- A) I cierta, II cierta
- B) I cierta, II falsa
- C) I falsa, II cierta
- D) I falsa, II falsa

20.- Señala si son o no ciertas las siguientes afirmaciones:

- I.- El modelo tradicional basado en herramientas de ETL, Datawarehouses y herramientas de Business Intelligence y Data Mining gestiona y analiza datos estructurados.
- II.- Los sistemas Datawarehouse se reemplazan por sistemas Big Data.

- A) I cierta, II cierta
- B) I cierta, II falsa
- C) I falsa, II cierta
- D) I falsa, II falsa