

PRIMER APELLIDO C	SEGUNDO APELLIDO M	NOMBRE DEL ALUMNO/A H	DÍA NACIMIENTO 30	MES NACIMIENTO 12
----------------------	-----------------------	--------------------------	----------------------	----------------------

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

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 CONTESTACIONES QUE NO ESTÉN EN LA TABLA.**
- Para contestar, marca la letra de la respuesta que consideres correcta en el recuadro adecuado.
- Se quiere 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 <u>C</u>	SEGUNDO APELLIDO <u>M</u>	NOMBRE DEL ALUMNO/A <u>M</u>	DÍA NACIMIENTO <u>30</u>	MES NACIMIENTO <u>12</u>
-----------------------------	------------------------------	---------------------------------	-----------------------------	-----------------------------

TABLA DE RESPUESTAS

PREGUNTA 1	A	<input checked="" type="radio"/> B	C	D
PREGUNTA 2	<input checked="" type="radio"/> A	B	C	D
PREGUNTA 3	<input checked="" type="radio"/> A	B	C	<input checked="" type="radio"/> D
PREGUNTA 4	A	<input checked="" type="radio"/> B	C	D
PREGUNTA 5	A	B	C	<input checked="" type="radio"/> D
PREGUNTA 6	A	B	C	<input checked="" type="radio"/> D
PREGUNTA 7	A	B	C	<input checked="" type="radio"/> D
PREGUNTA 8	<input checked="" type="radio"/> A	B	C	D
PREGUNTA 9	A	B	C	D
PREGUNTA 10	A	B	C	D

PREGUNTA 11	<input checked="" type="radio"/> A	B	C	D
PREGUNTA 12	<input checked="" type="radio"/> A	B	C	D
PREGUNTA 13	A	B	C	D
PREGUNTA 14	A	B	C	D
PREGUNTA 15	A	<input checked="" type="radio"/> B	C	D
PREGUNTA 16	A	<input checked="" type="radio"/> B	C	D
PREGUNTA 17	A	B	C	D
PREGUNTA 18	A	<input checked="" type="radio"/> B	C	D
PREGUNTA 19	A	B	C	<input checked="" type="radio"/> D
PREGUNTA 20	A	B	C	<input checked="" type="radio"/> D

PRIMER APELLIDO	SEGUNDO APELLIDO	NOMBRE DEL ALUMNO/A	DÍA NACIMIENTO	MES NACIMIENTO
C	H	H	30	12

- 1.- ¿Cuál es el componente que se utiliza para planificar, orquestar o automatizar flujos de trabajo en Hadoop?:
- A) Impala.
 - ☒ B) Oozie.
 - C) YARN.
 - D) Hive.
- 2.- ¿Cuál de los siguientes componentes de Hadoop permite acceder a ficheros de datos estructurados o semiestructurados en HDFS como si fueran una tabla de una base de datos relacional utilizando un lenguaje similar a SQL?
- ☒ A) Apache Hive
 - B) Apache Pig
 - C) Apache HBase
 - D) Apache Flume
- 3.- ¿Cuándo es adecuado utilizar Hadoop?
- ☒ a) Cuando se tiene requisitos de transaccionalidad muy estrictos.
 - b) Cuando los datos tienen un formato fijo y no cambian.
 - c) Cuando se necesita una escalabilidad baja.
 - ☒ d) Ninguna de las anteriores.
- 4.- ¿Cómo se llama el principal fichero de configuración para el servicio HDFS?:
- A) core-config.xml.
 - ☒ B) hdfs-site.xml.
 - C) hadoop-commons.xml.
 - D) hdfs-config.xml.
- 5.- ¿Qué comando HDFS se utiliza para copiar un archivo del sistema de archivos distribuido de Hadoop al sistema local?
- A) hadoop fs -ls
 - B) hadoop fs -mkdir
 - C) hadoop fs -put
 - ☒ D) hadoop fs -get
- 6.- ¿Cómo consigue HDFS tener tolerancia a fallos?
- A) Dividiendo los ficheros en bloques.
 - B) Almacenando los bloques en diferentes nodos.
 - C) Replicando los bloques en varios nodos.
 - ☒ D) Todas las anteriores.

shuffle, order, reduce

PRIMER APELLIDO	SEGUNDO APELLIDO	NOMBRE DEL ALUMNO/A	DÍA NACIMIENTO	MES NACIMIENTO
C	H	M	30	12

- ☒ 8.- ¿Cuál es el propósito de la fase Map en un proceso MapReduce?
- A) Dividir los datos en partes más pequeñas.
 - B) Aplicar una función a cada registro individualmente.
 - C) Ordenar los registros por clave. ✗
 - ☒ D) Combina los registros con la misma clave.
- ☒ 9.- ¿Cuál es el propósito de la fase Reduce en un proceso MapReduce?
- ☒ A) Dividir los datos en partes más pequeñas.
 - B) Aplicar una función a cada registro individualmente.
 - C) Ordenar los registros por clave. ✗
 - D) Procesar los registros con la misma clave.
- ☒ 9.- ¿Qué pasa si la salida de un mapper en MRJob se hace de la siguiente forma?
- yield None, value
- A) Se producirá un error.
 - B) La clave tendrá valor "value".
 - C) Todos los registros serán enviados al mismo reducer.
 - D) Los registros serán devueltos al mapper.
- ☒ 10.- ¿Qué sucede cuando se elimina una partición interna en Hive?
- A) La partición se elimina de la tabla, pero los datos permanecen en HDFS.
 - B) La partición y los datos asociados se eliminan de la tabla y de HDFS.
 - C) La partición se elimina de HDFS, pero los datos permanecen en la tabla.
 - D) La partición se mueve a un directorio temporal antes de ser eliminada. ~
- ☒ 11.- Señala si son o no ciertas las siguientes afirmaciones:
- I.- Las consultas de Hive se traducen a procesos MapReduce.
 - II.- Los datos en Hive 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
- ☒ 12.- ¿Qué comando se utiliza para cargar datos en Hive?
- ☒ A) LOAD
 - B) READ
 - C) SELECT
 - D) Ninguna de las anteriores

PRIMER APELLIDO	SEGUNDO APELLIDO	NOMBRE DEL ALUMNO/A	DÍA NACIMIENTO	MES NACIMIENTO
C	N	M	30	12

✗ 13.- Señala si son o no ciertas las siguientes afirmaciones sobre las siguientes sentencias de PIG:

```
A = LOAD 'student' USING PigStorage() AS (name:chararray, age:int, gpa:float);
B = FOREACH A GENERATE name;
DESCRIBE B;
```

- I.- Las sentencias no se habrán ejecutado al no haber incluido una sentencia DUMP.
- II.- No se puede hacer un DESCRIBE sin un DUMP previo.

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

✗ 14.- 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 un Dataframe de Spark.
- II.- Se puede procesar directamente con un RDD Spark.

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

✗ 15.- ¿Qué es un UDF en Apache Pig?

- A) Un formato de nombres de usuario.
- ✗ B) Una función definida por el usuario para procesar datos en Pig Latin.
- C) Una interfaz de usuario para consultar datos en Pig Latin.
- D) Un formato universal de disco.

PRIMER APELLIDO	SEGUNDO APELLIDO	NOMBRE DEL ALUMNO/A	DÍA NACIMIENTO	MES NACIMIENTO
<input type="text" value="C"/>	<input type="text" value="M"/>	<input type="text" value="M"/>	<input type="text" value="30"/>	<input type="text" value="12"/>

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

SELECT * FROM airports LIMIT 10

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

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

→ X 17.- Señala si son o no ciertas las siguientes afirmaciones sobre la siguiente consulta:

- I.- En "dataframes" de Spark "groupBy" es una transformación.
- II.- En "dataframes" de Spark "limit" es una transformación.

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

18.- ¿Qué lenguajes de programación se pueden usar en Databricks?

- A) Python, R, Scala, SQL y Java. ✗
- ☒ B) Python y SQL.
- C) Python, SQL y Java. ✓
- D) Python, SQL y C.

19.- ¿Cuál de las siguientes afirmaciones sobre EMR NO es correcta?

- A) EMR permite arrancar clústers Hadoop rápidamente, por lo que es muy útil para hacer pruebas con Hadoop.
- B) EMR puede adaptar el número de servidores a la carga real que esté soportando, por lo que sólo pagas por el uso real.
- C) EMR permite configurar qué componentes del ecosistema Hadoop arrancar.
- ☒ D) EMR, que son las siglas de Elastic MapReduce, sólo permite MapReduce como framework para procesar datos.

PRIMER APELLIDO	SEGUNDO APELLIDO	NOMBRE DEL ALUMNO/A	DÍA NACIMIENTO	MES NACIMIENTO
C	M	M	30	12

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

- I.- La implementación de un Data Lake es *más económica* menos económica que la de un Datawarehouse. *falsa*
- II.- Un Data Lake sólo almacena datos estructurados, mientras que un Datawarehouse puede almacenar cualquier tipo de dato. *→ al revés*

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

