

PRIMER APELLIDO <input type="text" value="C"/>	SEGUNDO APELLIDO <input type="text" value="M"/>	NOMBRE DEL ALUMNO/A <input type="text" value="D"/>	DÍA NACIMIENTO <input type="text" value="19"/>	MES NACIMIENTO <input type="text" value="01"/>
---	--	---	---	---

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 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	SEGUNDO APELLIDO	NOMBRE DEL ALUMNO/A	DÍA NACIMIENTO	MES NACIMIENTO
C	my	P	12	01

### TABLA DE RESPUESTAS

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

PRIMER APELLIDO	SEGUNDO APELLIDO	NOMBRE DEL ALUMNO/A	DÍA NACIMIENTO	MES NACIMIENTO
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	19	01

1.- ¿Cuál es el comando de HDFS para crear un directorio?:

- ☒ A) mkdir.
- B) create\_dir.
- C) rmdir.
- D) make\_dir.

2.- ¿Qué tipo de escalado es el más apropiado para Hadoop?:

- ☒ A) Escalado horizontal.
- B) Escalado vertical.
- C) Escalado tangencial.
- D) Hadoop no escala.

3.- ¿Qué componente de Hadoop permite importar o exportar datos estructurados desde bases de datos relacionales a Hadoop y viceversa?

- ☒ A) Apache Sqoop.
- B) Apache Hive.
- C) Apache Pig.
- D) Apache HBase.

4.- ¿Por qué los bloques en HDFS tienen un tamaño tan grande?

- A) Para mejorar el rendimiento en operaciones pequeñas.
- B) Para permitir el acceso concurrente de múltiples clientes.
- C) Para reducir la cantidad de bloques necesarios en ficheros grandes.
- ☒ D) Para facilitar la replicación de los bloques en diferentes nodos.

5.- ¿Qué es el nodo NameNode en HDFS?

- ☒ A) Un nodo de respaldo para el NameNode principal.
- ☒ B) El nodo maestro que mantiene la metainformación del sistema de archivos HDFS.
- C) Un nodo worker que almacena los bloques de los ficheros.
- D) Un nodo que toma puntos de control de los metadatos del sistema de ficheros del Namenode.

6.- ¿Qué componente del ecosistema Hadoop permite acceder a los ficheros de HDFS con lenguaje SQL, como si fueran tablas de una base de datos?:

- A) MapReduce.
- ☒ B) Hive.
- C) HDFS.
- D) YARN.



PRIMER APELLIDO	SEGUNDO APELLIDO	NOMBRE DEL ALUMNO/A	DÍA NACIMIENTO	MES NACIMIENTO
C	M	D	19	01

7.- ¿Cuál es el papel del ApplicationMaster en YARN?

- A) Negociar con el ResourceManager los recursos necesarios para la ejecución de las tareas de su aplicación.
- ☒ B) Controlar la ejecución de todas las tareas en YARN.
- C) Proporcionar información sobre el consumo de recursos en los nodos worker.
- D) Supervisar el ciclo de vida de los contenedores de aplicaciones.

8.- ¿Cuáles son las fases de ejecución de un trabajo MapReduce?

- A) map, shuffle, reduce, order *m s n o*
- B) map, order, shuffle, reduce *m o s n*
- ☒ C) map, shuffle, order, reduce *m s o n*
- ☒ D) map, reduce, shuffle, order *m n s o*

9.- ¿Qué es el servicio NodeManager en YARN?

- A) El maestro que controla la ejecución de todas las tareas que están en ejecución.
- ☒ B) El servicio que recibe las peticiones de ejecución por parte de los clientes.
- C) El servicio que se ejecuta en cada nodo worker y supervisa la ejecución de las tareas en contenedores.
- D) El servicio que asigna prioridades y establece los recursos/containers que disfrutará cada aplicación.

10.- ¿Qué línea de código se utiliza para especificar la entrada de datos en MRJob si no se definen etapas ("steps")?

- A) def combiner(self, input\_path):
- B) def reducer(self, key, values):
- C) def input\_data(self, input\_path):
- ☒ D) def mapper(self, key, value):

11.- ¿Qué comando se utiliza para cargar datos en Pig Latin?

- A) IMPORT
- B) READ
- ☒ C) LOAD
- D) SELECT

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

- I.- Antes de cargar los datos en Pig hay que definir el esquema.
- II.- Antes de cargar los datos en Hive hay que definir el esquema.

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

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="D"/>	<input type="text" value="19"/>	<input type="text" value="6"/>

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

- I.- En MRJob no es obligatorio definir un mapper.
- II.- En MRJob no puede haber más de un mapper.

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

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

- I.- Un fichero "dbc" al abrirlo en Databricks puede contener directorios y ficheros.
- II.- Un fichero "ipynb" al abrirlo con Jupyter puede contener directorios y ficheros.

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

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

- I.- La implementación de un Data Lake es más costosa que la de un Datawarehouse.
- II.- Un Datawarehouse sólo almacena datos estructurados, mientras que un Data Lake puede almacenar cualquier tipo de dato.

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

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

- I.- Una acción sobre un "dataframe" de Spark devuelve otro "dataframe".
- II.- Una transformación sobre un "dataframe" de Spark devuelve otro "dataframe".

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

PRIMER APELLIDO	SEGUNDO APELLIDO	NOMBRE DEL ALUMNO/A	DÍA NACIMIENTO	MES NACIMIENTO
C	M	P	19	01

17.- ¿Qué tipo de celda NO es válida en un fichero "dbc" de Databricks:

- ☒ A) Java
- ☐ B) Markdown
- ☐ C) Python
- ☐ D) Scala

18.- ¿Cuáles de las siguientes herramientas permiten analizar datos en tiempo real?

- ☒ A) Apache Storm.
- ☒ B) Apache Flink.
- ☒ C) Apache Spark.
- ☒ D) Todas las anteriores.

19.- ¿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.

20.- ¿Cuál de las siguientes afirmaciones NO es correcta en relación con los Data Lakes?

- ☐ A) Hadoop es una buena plataforma para implementar un Data Lake.
- ☐ B) Un Data Lake es más fácil de gestionar que un Datawarehouse.
- ☒ C) Los Data Lakes intentan ser un repositorio de datos único para toda la empresa.
- ☐ D) Frente a los Datawarehouses tradicionales, un Data Lake ofrece más funcionalidad.