

Universidad Nacional Abierta y a Distancia
Vicerrectoría Académica y de Investigación
Curso: Big Data
Código: 202016911

Guía de actividades y rúbrica de evaluación – Tarea 4
 Almacenamiento y Consultas de Datos en Big Data

1. Descripción de la actividad

Tipo de actividad: En grupo colaborativo	
Momento de la evaluación: Intermedio	
Puntaje máximo de la actividad: 120 puntos	
La actividad inicia el: miércoles, 30 de octubre de 2024	La actividad finaliza el: martes, 26 de noviembre de 2024
Con esta actividad se espera conseguir los siguientes resultados de aprendizaje: Resultado de aprendizaje 3: Aplicar técnicas de análisis y visualización a grandes conjuntos de datos para obtener información útil. Resultado de aprendizaje 4: Explorar las últimas tendencias en Big Data, como la computación en la nube, el IoT, el aprendizaje automático y Bases de datos NoSQL.	
La actividad consiste en: Fase 1: MongoDB 1. Diseño de la base de datos: Seleccionar un caso de uso donde sea apropiado utilizar una base de datos NoSQL como MongoDB (ej: almacenamiento de datos de redes sociales, catálogos de productos, logs de aplicaciones). Diseñar el esquema de la base de datos en MongoDB, definiendo las colecciones, documentos y campos. 2. Implementación en MongoDB: <ul style="list-style-type: none"> ○ Crear una base de datos en MongoDB e insertar un conjunto de datos de prueba (al menos 100 documentos). ○ Implementar consultas utilizando el lenguaje de consulta de MongoDB para realizar las siguientes operaciones: 	

1. Consultas básicas (inserción, selección, actualización y eliminación de documentos).
2. Consultas con filtros y operadores.
3. Consultas de agregación para calcular estadísticas (contar, sumar, promediar, etc.).

3. **Documentación:**

- Documentar el diseño de la base de datos en MongoDB.
- Explicar el código de las consultas y los resultados obtenidos.

Fase 2: Apache HBase

1. **Creación de la tabla en HBase:** Utilizar un conjunto de datos de gran volumen (puede ser el mismo que en la Fase 1 o uno diferente) y definir una tabla en HBase, especificando las familias de columnas y configuraciones necesarias.

2. **Operaciones en HBase:** Implementar operaciones en Hbase para:

- Cargar los datos en la tabla de HBase.
- Realizar consultas de selección, filtrado y recorrido sobre los datos.
- Realizar operaciones de escritura (inserción, actualización y eliminación) sobre los datos.

3. **Documentación:**

- Describir la estructura de la tabla en HBase.
- Explicar las operaciones realizadas en HBase y los resultados obtenidos.

Fase 3: Investigación: Realizar un ensayo comparando los diferentes tipos de bases de datos NoSQL (Clave-valor, documentos, columnas, grafos). Discuta las ventajas e inconvenientes de cada tipo, así como sus casos de uso más apropiados.

Entregables:

1. **Foro de discusión:**

- Compartir el caso de uso seleccionado para MongoDB.
- Publicar avances y dudas durante el desarrollo de la actividad.

2. **Repositorio de código:** Incluir el código de las consultas en MongoDB y HBase.
3. **Informe académico (Tarea4_grupo.pdf):** El documento debe contener:
 - Portada
 - Introducción
 - Objetivos
 - **Fase 1: MongoDB**
 1. Descripción del caso de uso y diseño de la base de datos.
 2. Código de las consultas y análisis de resultados.
 - **Fase 2: Apache HBase**
 1. Descripción del conjunto de datos y estructura de la tabla en HBase.
 2. Código de las operaciones en HBase y análisis de resultados.
 - **Fase 3:** Respuesta a la actividad de investigación.
 - Enlace al repositorio de código.
 - 5 conclusiones generales del grupo.
 - Referencias Bibliográficas (Normas APA).

Para el desarrollo de la actividad tenga en cuenta que:

En el entorno de Información inicial debe:

Revisar la agenda del curso y ajustar sus actividades para realizar la entrega del producto final de acuerdo con la fecha establecida.

En el entorno de Aprendizaje debe:

- Realizar las lecturas recomendadas para cada tema y participar activamente en las actividades propuestas.
- Participar en el foro de la tarea 4 – Almacenamiento y Consultas de Datos en Big Data, publicando los avances de la actividad y el informe académico
- Revisar el trabajo final para que esté en conformidad con lo solicitado en la guía y rúbrica de la actividad.

En el entorno de Evaluación debe:

- Realizar la entrega del documento final en formato PDF.

Evidencias de trabajo independiente:

Las evidencias de trabajo independiente para entregar son:

- **Participación en el foro:** Al menos 2 aportes de calidad.
- **Implementación de consultas:** Implementar las consultas en MongoDB y HBase de forma individual.
- **Conclusiones individuales:** Redactar dos (2) conclusiones individuales sobre la Unidad 3.

Nota: Recuerde que su trabajo será evaluado con la herramienta turnitin, por lo tanto, es muy importante citar correctamente las referencias en el ensayo.

Evidencias de trabajo grupal:

Las evidencias de trabajo grupal a entregar son:

1. **Informe académico:** Integrar el trabajo individual en un único informe coherente y bien estructurado
2. **Entrega del documento final en formato PDF (Tarea4_grupo.pdf):** Un solo documento por grupo con la información completa. El documento debe contener:
 - Portada
 - Introducción
 - Objetivos
 - **Fase 1: MongoDB**
 1. Descripción del caso de uso y diseño de la base de datos.
 2. Código de las consultas y análisis de resultados.
 - **Fase 2: Apache HBase**
 1. Descripción del conjunto de datos y estructura de la tabla en HBase.
 2. Código de las operaciones en HBase y análisis de resultados.
 - **Fase 3:** Respuesta a la actividad de investigación.
 - Enlace al repositorio de código.
 - 5 conclusiones generales del grupo.

- Referencias Bibliográficas (Normas APA).

Aspectos a tener en cuenta:

- 1. Investigación:** Investigar a profundidad las características, ventajas y desventajas de MongoDB y HBase.
- 2. Funcionalidad:** Las consultas implementadas deben ser correctas y eficientes.
- 3. Análisis de resultados:** Interpretar y analizar los resultados de las consultas, extrayendo conclusiones relevantes

Formato:

- 1.** Seguir las convenciones de código para MongoDB y HBase.
- 2.** Utilizar un lenguaje claro y conciso en la documentación.

2. Lineamientos generales para la elaboración de las evidencias de aprendizaje a entregar.

Para evidencias elaboradas **en grupo colaborativamente**, tenga en cuenta las siguientes orientaciones

1. Realice un reconocimiento general del curso y de cada uno de los entornos antes de abordar el desarrollo de las actividades.
2. Identifique y lea los recursos y los referentes de la unidad 1 que corresponden a la actividad.
3. Intervenga en el foro de discusión aplicando las normas de Netiqueta Virtual, evidenciando siempre respeto por las ideas de sus compañeros y del cuerpo docente.
4. Antes de entregar el producto solicitado revise que cumpla con todos los requerimientos que se señalaron en esta guía de actividades, rúbrica de evaluación y por parte del tutor en el foro de discusión.
5. No cometa fraudes, ni plagios ni actos que atenten contra el normal desarrollo académico de las actividades.
6. Todos los integrantes del grupo deben participar con sus aportes en el desarrollo de la actividad.
7. En cada grupo deben elegir un solo integrante que se encargará de entregar el producto solicitado en el entorno que haya señalado el docente.
8. Solo se deben incluir como autores del producto entregado, a los integrantes del grupo que hayan participado con aportes durante el tiempo destinado para la actividad.

Tenga en cuenta que todos los productos escritos individuales o grupales deben cumplir con las normas de ortografía y con las condiciones de presentación que se hayan definido.

En cuanto al uso de referencias considere que el producto de esta actividad debe cumplir con las normas **APA**

En cualquier caso, cumpla con las normas de referenciación y evite el plagio académico, para ello puede apoyarse revisando sus productos escritos mediante la herramienta Turnitin que encuentra en el campus virtual.

Considere que en el acuerdo 029 del 13 de diciembre de 2013, artículo 99, se considera como faltas que atentan contra el orden académico, entre otras, las siguientes: literal e) "El plagiar, es decir, presentar como de su propia autoría la totalidad o parte de una obra, trabajo, documento o invención realizado por otra persona. Implica también el uso de citas o referencias faltas, o proponer citad donde no haya coincidencia entre ella y la referencia" y liberal f) "El reproducir, o copiar con fines de lucro, materiales educativos o resultados de productos de investigación, que cuentan con derechos intelectuales reservados para la Universidad"

Las sanciones académicas a las que se enfrentará el estudiante son las siguientes:

- a) En los casos de fraude académico demostrado en el trabajo académico o evaluación respectiva, la calificación que se impondrá será de cero puntos sin perjuicio de la sanción disciplinaria correspondiente.
- b) En los casos relacionados con plagio demostrado en el trabajo académico cualquiera sea su naturaleza, la calificación que se impondrá será de cero puntos, sin perjuicio de la sanción disciplinaria correspondiente.

3. Formato de Rúbrica de evaluación

Tipo de actividad: En grupo colaborativo	
Momento de la evaluación: Intermedio	
La máxima puntuación posible es de 120 puntos	
<p>Primer criterio de evaluación:</p> <p>Cumple con el diseño e Implementación de Soluciones</p> <p>Este criterio representa 40 puntos del total de 120 puntos de la actividad.</p>	<p>Nivel alto: Se diseña e implementa correctamente la base de datos en MongoDB y la tabla en HBase, aplicando buenas prácticas y optimizando el almacenamiento y las consultas. Se realizan consultas complejas que demuestran un buen entendimiento de ambas tecnologías. Si su trabajo se encuentra en este nivel puede obtener entre 21 puntos y 40 puntos</p> <p>Nivel Medio: Se presentan errores u omisiones en el diseño o implementación de las soluciones. Las consultas son básicas o no se aprovechan las funcionalidades avanzadas de MongoDB o HBase. Si su trabajo se encuentra en este nivel puede obtener entre 6 puntos y 20 puntos</p> <p>Nivel bajo: El diseño o implementación de las soluciones es incorrecto o incompleto. Las consultas son erróneas o no se ejecutan. Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 5 puntos</p>
<p>Segundo criterio de evaluación:</p> <p>Realiza el análisis y comparación entre MongoDB y HBase</p> <p>Este criterio representa 40 puntos del total de 120 puntos de la actividad</p>	<p>Nivel alto: Se analizan e interpretan correctamente los resultados de las consultas, extrayendo conclusiones relevantes. Se realiza una comparación detallada y bien argumentada entre MongoDB y HBase, destacando las ventajas e inconvenientes de cada tecnología para el caso de uso seleccionado. Si su trabajo se encuentra en este nivel puede obtener entre 21 puntos y 40 puntos</p> <p>Nivel Medio: El análisis de resultados es superficial o poco preciso. La comparación entre MongoDB y HBase es incompleta o poco profunda. Si su trabajo se encuentra en este nivel puede obtener entre 6 puntos y 20 puntos</p> <p>Nivel bajo: No se realiza un análisis adecuado de resultados. La comparación entre tecnologías es inexistente o errónea.</p>

	Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 5 puntos
Tercer criterio de evaluación: Elabora el informe académico Este criterio representa 40 puntos del total de 120 puntos de la actividad	Nivel alto: El informe escrito es claro, conciso, bien organizado y cumple con los requisitos de formato. El código está bien documentado. Se evidencia un trabajo en equipo equitativo y de alta calidad. Si su trabajo se encuentra en este nivel puede obtener entre 21 puntos y 40 puntos Nivel Medio: El informe presenta errores de formato, falta de claridad o se evidencia una participación desigual en el grupo. Si su trabajo se encuentra en este nivel puede obtener entre 6 puntos y 20 puntos Nivel bajo: El informe es deficiente, desorganizado o se evidencia falta de trabajo en equipo. Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 5 puntos