

Tema 5: Excel para Big Data

1. Introducción a Excel para Big Data.

- Realiza una breve explicación sobre cómo Excel puede ser utilizado para trabajar con conjuntos de datos más grandes y complejos.

Excel se ha convertido en una herramienta útil para algunas tareas de Big Data gracias a la incorporación de complementos como Power Query y Power Pivot, así como a mejoras en su capacidad para manejar volúmenes masivos de datos (hasta 1¹ millón de filas por hoja) y amplias capacidades de procesamiento en las últimas versiones (fórmulas mejoradas, mejor uso de la memoria y la optimización de la CPU²).

2. Manejo de grandes conjuntos de datos.

- Indica entre 3 y 5 consejos para trabajar de manera eficiente con grandes volúmenes de datos en Excel, como la optimización de fórmulas y el uso de funciones de tabla.

Optimización de fórmulas: las fórmulas simples reducen la carga de trabajo, lo que mejora el rendimiento de Excel, especialmente cuando se trabaja con grandes conjuntos de datos.

Uso de tablas: las tablas se actualizan automáticamente cuando se modifican los datos, lo que elimina la necesidad de ajustar las fórmulas manualmente. Asimismo, facilitan la organización y depuración de datos.

Segmentación de datos: Distribuir los datos en diferentes hojas o archivos hace que Excel sea más fluido y simplifica la gestión de la información.

Power Query y Power Pivot: estas son dos herramientas se recomiendan para analizar grandes volúmenes de datos. Ambas mejoran las capacidades de análisis en Excel.

3. Herramientas de análisis para Big Data.

- Menciona herramientas como Power Query y Power Pivot que permiten importar, transformar y analizar grandes conjuntos de datos de manera más eficiente en Excel.

Power Query es una herramienta que permite importar, limpiar y transformar datos con gran facilidad. Permite conectarse a una gran variedad de fuentes, como archivos Excel, CSV, SQL, TXT y páginas web, lo que facilita mucho la realización de tareas como filtrar, combinar y limpiar datos, sin importar la fuente de nuestros datos. El proceso de preparación de datos se puede automatizar y además, Power Query ofrece la función de registro de pasos, mostrándote en todo momento el flujo de transformaciones al que han sido sometidos los datos. Es un complemento de Excel, para gestionar grandes conjuntos de datos.

¹ Microsoft. (2023). What to do if a dataset is too large for the Excel grid. Microsoft Office Support. <https://support.microsoft.com/en-gb/office/what-to-do-if-a-data-set-is-too-large-for-the-excel-grid-976e6a34-9756-48f4-828c-ca80b3d0e15c>

² Microsoft. (2023). Excel performance and limit improvements. Microsoft Learn. <https://learn.microsoft.com/en-us/office/vba/excel/concepts/excel-performance/excel-performance-and-limit-improvements>

Por otro lado, encontramos, Power Pivot, otro complemento, que utiliza DAX para crear relaciones entre tablas y nos ofrece herramientas para análisis más complejos, brindándonos capacidades de modelado de datos y análisis avanzados en Excel. Es una herramienta muy útil, también, si necesitamos manejar grandes volúmenes de datos.

La combinación de Power Query y Power Pivot en Excel, nos permite analizar gran cantidad de datos de una forma mejorada y más eficaz.

4. Limitaciones y consideraciones.

- Discute las limitaciones de Excel en comparación con herramientas específicas de Big Data, como la capacidad de procesamiento y la escalabilidad.

Excel tiene límites en comparación con ciertas soluciones de Big Data, a pesar de su evolución para manejar mayores volúmenes de datos:

- Procesamiento de datos masivo: Excel puede ser lento, especialmente en sistemas de bajos recursos, cuando se trata de conjuntos de datos muy grandes o complejos.
- Escalabilidad: Excel tiene un número máximo de filas y columnas, lo que puede ser una barrera para algunos análisis, mientras que las herramientas especializadas de Big Data pueden manejar terabytes de datos o más.
- Funcionalidades para análisis avanzado: si bien Excel es una gran herramienta, algunas tecnologías de Big Data brindan aprendizaje automático, análisis predictivo y procesamiento de datos en tiempo real que Excel no ofrece.

5. Recursos adicionales y conclusiones.

- Proporciona enlaces a recursos adicionales donde los usuarios puedan aprender más sobre cómo utilizar Excel para manejar y analizar Big Data de manera más avanzada. También, realiza una breve conclusión de los principales aprendizajes del módulo.

Para obtener más información sobre cómo utilizar Excel para manejar y analizar Big Data, se puede consultar los siguientes recursos:

Recurso	Link
Documentación de Excel	https://support.microsoft.com/en-us/excel
Learn Microsoft	https://learn.microsoft.com/en-us/training/modules/modern-analytics-intro/
Codecademy	https://www.codecademy.com/learn/analyze-data-with-microsoft-excel
LinkedIn Training	https://www.linkedin.com/learning/learning-excel-data-analysis-18868618

Tras terminar el módulo de Excel para análisis de datos, he mejorado mis conocimientos de esta herramienta. He descubierto cómo Microsoft Excel es una herramienta poderosa no solo para tareas diarias, sino también para un análisis de datos.

He aprendido a usar la interfaz de Excel y sus muchas funciones para organizar, buscar y filtrar datos. He usado fórmulas avanzadas y funciones de búsqueda y lógica, lo que me permite desentrañar y examinar los datos de maneras que antes no conocía.

El uso de tablas y tablas dinámicas ha enriquecido mi habilidad para consolidar y analizar cantidades significativas de datos con eficacia. Herramientas adicionales de análisis en Excel han ampliado mi capacidad para proyectar tendencias y solucionar problemas complejos. Además, la habilidad para visualizar datos con gráficos y tableros ha sido fundamental para mostrar y comunicar hallazgos de manera fácil y comprensible para otros.

Un avance notable en mi aprendizaje ha sido el descubrimiento de Power Query y Power Pivot. Power Query me ha permitido importar, limpiar y preparar datos de fuentes diversas, mientras que conocer la existencia de Power Pivot, me facilitará desde ahora, la creación de modelos de datos relacionales complejos.

En resumen, este nuevo conocimiento adquirido, cimienta los fundamentos que necesito para seguir desarrollándome como analista de datos.