

Título de la tarea: Introducción a Hadoop

Curso de Especialización y módulo: Inteligencia Artificial y Big Data - Big Data Aplicado

¿Qué contenidos o resultados de aprendizaje trabajaremos?

Resultados de aprendizaje

- ✓ **RA1.** Gestiona sistemas de almacenamiento y el amplio ecosistema alrededor de ellos facilitando el procesamiento de grandes cantidades de datos sin fallos y de forma rápida.

Contenidos

- 1.- Motivación y origen.
- 2.- Apache Hadoop a alto nivel.
 - 2.1.- ¿Qué es Apache Hadoop?
 - 2.2.- Ecosistema Hadoop y distribuciones.
 - 2.3.- Arquitectura.
 - 2.4.- Beneficios, desventajas y dificultades.
- 3.- Instalación de Apache Hadoop con Docker.

Siguiente »

1.- Descripción de la tarea.



Instalación del entorno Hadoop en un equipo local

En esta tarea instalarás en tu equipo informático el entorno de Hadoop con el que realizarás el resto de prácticas del módulo.

La instalación se hará en contenedores de Docker. En próximas prácticas usarás el terminal y Python para escribir tus programas. En esta tarea desplegarás un servidor de Jupyter en el que podrás escribir tu código.



[Lukas Bieri](#) (Dominio público)

¿Qué te pedimos que hagas?

Para esta tarea vas a necesitar un equipo con una capacidad de procesamiento equivalente a un microprocesador i5 o superior y una memoria RAM de un mínimo de 8GB. Se recomienda que el sistema operativo anfitrión sea Linux, aunque también se podrá realizar sobre Windows o MAC:

Descárgate el enlace al [fichero zip](#) inferior, descomprímelo y sigue las instrucciones que se dieron en la guía práctica. En resumen lo que hay que hacer es situar el libro de Jupyter (el que tiene extensión "ipynb") y el directorio de imágenes (llamado "img") en el contenedor "namenode" para que sea accesible por Jupyter. Después abres el fichero en el navegador (en el servidor de Jupyter) y sigues las instrucciones de realización.

Para entregar debes subir un fichero comprimido que contenga el fichero "ipynb" con las celdas ejecutadas junto con las imágenes que hayas utilizado. En la guía práctica se explica con un ejemplo como hacer esto.

NOTA IMPORTANTE

Para todas las preguntas es necesario entregar un documento en el que aparezca la pregunta junto con tu respuesta razonada. Incluye cualquier diagrama que consideres interesante para explicar la solución que has desarrollado.

2.- Información de interés.



Recursos necesarios y recomendaciones

Recursos necesarios

- ✓ Conexión a Internet para consultar la información o investigar sobre las cuestiones planteadas.
- ✓ Navegador web.
- ✓ Equipo informático con procesador equivalente a un i5 o superior y un mínimo de 8GB de RAM.
- ✓ Recomendable SO Linux. No te preocupes si no lo tienes instalado, también puedes hacer la práctica con Windows o MAC.

Recomendaciones

- ✓ Antes de abordar la tarea:
 - lee con detenimiento la unidad, consulta los enlaces para saber más, revisa la guía práctica y realiza paso a paso lo que allí se indica.
 - Si tienes problemas, escribe en el foro.
 - Realiza el examen online de la unidad, y consulta nuevamente las dudas que te surjan. Solo cuando lo tengas todo claro, debes abordar la realización de la tarea.
- ✓ No olvides crear un "zip" que contenga el archivo "ipynb" con las celdas ejecutadas y las imágenes que hayas utilizado.



Indicaciones de entrega

Una vez realizada la tarea, el envío se realizará a través de la plataforma. El archivo se nombrará siguiendo las siguientes pautas:

Apellido1_Apellido2_Nombre_BDA01_Tarea

Asegúrate que el nombre no contenga la letra ñ, tildes ni caracteres especiales extraños. Así por ejemplo la alumna **Begoña Sánchez Mañas para la primera unidad del MP de BDA01**, debería nombrar esta tarea como...

sanchez_manas_begona_BDA01_Tarea

« Anterior Siguiente »

3.- Evaluación de la tarea.



Criterios de evaluación implicados

Criterios de evaluación RA1

- ✓ a) Se ha determinado la importancia de los sistemas de almacenamiento para depositar y procesar grandes cantidades de cualquier tipo de datos rápidamente.
- ✓ b) Se ha comprobado el poder de procesamiento de su modelo de computación distribuida.
- ✓ c) Se ha probado la tolerancia a fallos de los sistemas.
- ✓ d) Se ha determinado que se pueden almacenar tantos datos como se desee y decidir cómo utilizarlos más tarde.
- ✓ e) Se ha visualizado que el sistema puede crecer fácilmente añadiendo módulos.

¿Cómo valoramos y puntuamos tu tarea?

Rúbrica de la tarea	
Pregunta 1	2 puntos
Pregunta 2	2 puntos
Pregunta 3	2 puntoss
Pregunta 4	2 puntos
Pregunta 5	2 puntos
Pregunta 6 (opcional)	2 puntos
La puntuación máxima será de 10 puntos.	
Redacción clara y correcta, sin errores ortográficos	Se resta 0,1 puntos por cada error ortográfico o

expresiones
incorrectas.

« Anterior