

# SBD02.- Almacenamiento de Datos.

## Orientaciones Alumnado

En esta unidad de trabajo vamos a descubrir la importancia de ser capaces de almacenar datos de forma distribuida como ingrediente fundamental para nuestros sistemas Big Data.

Ello incluye saber qué es un sistema de ficheros y qué es sistema de ficheros distribuido.

También descubriremos qué son las bases de datos NoSQL y sus principales tipos, para finalizar dando nuestros primeros pasos con MongoDB, una de las más bases de datos NoSQL más utilizadas.

### Datos generales de la Unidad de Trabajo

<b>Nombre completo del MP</b>	Sistemas de Big Data.	<b>Siglas MP</b>	SBD
<b>Nº y título de la UT</b>	02.- Almacenamiento de Datos.		
<b>Índice o tabla de contenidos</b>	<ul style="list-style-type: none"> <li>1.- Sistemas de Ficheros.               <ul style="list-style-type: none"> <li>1.1.- RAID.</li> <li>1.2.- Sistemas de ficheros distribuidos.                   <ul style="list-style-type: none"> <li>1.2.1.- Almacenamiento distribuido en memoria.</li> </ul> </li> <li>1.3.- Hadoop Distributed File System (HDFS).                   <ul style="list-style-type: none"> <li>1.3.1.- Características y funcionamiento.</li> <li>1.3.2.- Acceso mediante línea de comando.</li> <li>1.3.3.- Acceso mediante Python.</li> </ul> </li> </ul> </li> <li>2.- Bases de datos NoSQL.               <ul style="list-style-type: none"> <li>2.1.- Conceptos generales.                   <ul style="list-style-type: none"> <li>2.1.1.- Sharding.</li> <li>2.1.2.- Replicación.</li> <li>2.1.3.- Sharding con Replicación.</li> </ul> </li> <li>2.2.- Tipos de NoSQL.                   <ul style="list-style-type: none"> <li>2.2.1.- Documentales.</li> <li>2.2.2.- Clave-Valor.</li> <li>2.2.3.- Columnares.</li> <li>2.2.4.- Orientadas a Grafo.</li> </ul> </li> </ul> </li> <li>3.- Primeros pasos con MongoDB.</li> </ul>		
<b>Objetivos</b>	En esta unidad de trabajo vamos a descubrir la importancia de ser capaces de almacenar datos de forma distribuida como ingrediente fundamental para nuestros sistemas Big Data.		

	<p>Comenzamos viendo qué es un sistema de ficheros, para después descubrir la existencia de sistemas de almacenamiento que funcionan de modo distribuido entre nodos de un clúster y de sistemas que de forma específica realizan esa tarea no con unidades de almacenamiento sino en memoria.</p> <p>Más adelante conoceremos HDFS, el sistema de almacenamiento distribuido de Hadoop.</p> <p>A continuación haremos un repaso por el mundo de las bases de datos NoSQL, descubriendo tanto algunos conceptos básicos en relación a las mismas como qué tipos existen.</p> <p>Por último seguiremos el tutorial oficial de MongoDB, la cual es una de las bases de datos NoSQL más conocidas y utilizadas.</p>	
<b>Temporalización (estimación)</b>	<b>Tiempo necesario para estudiar los contenidos (h)</b>	23
	<b>Tiempo necesario para completar la tarea (h)</b>	1
	<b>Tiempo necesario para completar el examen (h)</b>	1
	<b>Nº de días que se recomienda dedicar a esta unidad</b>	25
	La temporalización anterior no deja de ser una estimación media, ya que el tiempo a invertir va a depender mucho de las circunstancias personales de cada cual.	
<b>Consejos y recomendaciones</b>	Es aconsejable entender bien todos los conceptos puesto que en las unidades sucesivas se hará referencia a ellos.	