

**1. DATOS INFORMATIVOS**

<b>DOMINIO:</b>	Hábitat, Infraestructura y Movilidad		
<b>CARRERA:</b>	Maestría virtual en Sistemas de Información mención Data Science		
<b>Asignatura/Módulo:</b>	Sistemas de Big Data		
<b>Paralelo:</b>			<b>N° horas 144</b>
<b>Plan de estudios:</b>	2024-02		H. aprendizaje en contacto con el docente: 16
<b>Prerrequisitos:</b>			
<b>Periodo académico:</b>	2024-02		H. aprendizaje autónomo: 120 H. aprendizaje práctico-experimental: 8
<b>Docente o Co-Docente 1:</b>	<b>Grado académico y título profesional:</b>		
Melgarejo Heredia Rafael	Doctor of Philosophy in Web Science Ingeniero en Sistemas Informáticos y de Computación		
<b>Breve reseña de la actividad</b>			
Docente PUCE desde 1996, principal desde 2003. Consultor informático 1994-2010, Perito Informático 2000-2013			
<b>Indicación de horario de atención al estudiante:</b>			
<b>Tutoría presencial:</b>	<b>Teléfono:</b>		
<b>Tutoría virtual:</b>	<b>Correo electrónico:</b>	<a href="mailto:nombre@puce.edu.ec">nombre@puce.edu.ec</a>	

**2. DESCRIPCIÓN DE LA ASIGNATURA**

**DESCRIPCIÓN DE COMPETENCIAS Y RESULTADOS DE APRENDIZAJE:**

<b>COMPETENCIAS TRANSVERSALES</b>	<b>COMPETENCIAS DISCIPLINARES DEL DOMINIO</b>	<b>RESULTADOS DE APRENDIZAJE DE LA CARRERA</b>	<b>RESULTADOS DE APRENDIZAJE DE LA ASIGNATURA</b>
1. Humanista y con proyecto vital	Evaluación del entorno complejo  Transformación pertinente y sostenible del hábitat	1. Recopilar datos de diversas fuentes de información, alineado con el objetivo específico planteado o el problema a resolver.	Utilizar Tecnologías y Herramientas de Big Data para grandes volúmenes de datos.
2. Comprometido social, política y ambientalmente	Innovación responsable con el medio ambiente  Transformación pertinente y sostenible del hábitat	2. Analizar datos estructurados y no estructurados utilizando técnicas y métodos estadísticos, asegurando el rigor y la precisión del análisis mediante el método científico	Aplicar Técnicas de Big Data en grandes volúmenes de datos.
3. Crítico y analítico	Evaluación del entorno complejo  Transformación pertinente y sostenible del hábitat	3. Desarrollar modelos de Inteligencia Artificial (IA) inductivos (predictivos) o deductivos (descriptivos), que apoyen a las instituciones en procesos de toma de decisiones estratégicas.	Evaluar los fundamentos de Big Data para el análisis de grandes volúmenes de datos
4. Creativo, innovador y emprendedor	Innovación responsable con el medio ambiente  Aplicación de lenguajes espaciales, objetuales, visuales y verbales	4. Crear visualizaciones de datos que faciliten la comprensión y apoyen la toma de decisiones	
5. Motivado a ser más	Evaluación del entorno complejo  Transformación pertinente y sostenible del hábitat	5. Aplicar principios éticos en la gestión y análisis de datos, con responsabilidad social empresarial.	
6. Comunicador asertivo	Evaluación del entorno complejo  Aplicación de lenguajes espaciales, objetuales, visuales y verbales		

#### 4. EVALUACIÓN DE LOGROS DE APRENDIZAJE

Resultado de aprendizaje	Definición del criterio de evaluación del RdA	Ponderación (en porcentaje)
Evaluar los fundamentos de Big Data para el análisis de grandes volúmenes de datos	Identificar y describir las características y componentes clave de Big Data	30
	Analizar casos de uso de Big Data en diversas industrias	30
	Evaluar la relevancia y el impacto de Big Data en la toma de decisiones empresariales.	40
Utilizar Tecnologías y Herramientas de Big Data para grandes volúmenes de datos.	Implementar soluciones de almacenamiento y procesamiento de Big Data utilizando tecnologías como Hadoop y Spark	30
	Desarrollar y ejecutar scripts y programas para la manipulación de grandes conjuntos de datos.	30
	Utilizar bases de datos NoSQL para el almacenamiento y consulta de datos no estructurados	40
Aplicar Técnicas de Big Data en grandes volúmenes de datos.	Realizar análisis exploratorios de grandes conjuntos de datos para identificar patrones y tendencias	30
	Desarrollar modelos de aprendizaje automático para predicción y clasificación utilizando herramientas como Python o R	30
	Interpretar y comunicar los resultados del análisis de Big Data a audiencias técnicas y no técnicas	40

**5. METODOLOGÍA**

Aprendizaje colaborativo, Aprendizaje basado en problemas

**6. RELACIÓN RESULTADOS DE APRENDIZAJE, EXPERIENCIAS DE APRENDIZAJES Y DIMENSIÓN DEL CONOCIMIENTO**

RESULTADOS DE APRENDIZAJE DE LA ASIGNATURA	Semana	DIMENSIÓN DEL CONOCIMIENTO (conceptos, hecho, procedimientos o principios)
<p>Evaluar los fundamentos de Big Data para el análisis de grandes volúmenes de datos</p>	1	<p>1. Fundamentos de BigData                      1.1. Introducción a Big Data, definición y características                      1.2. Describir las características y componentes clave de Big Data</p>
	2	<p>2. Aplicaciones de BigData                      2.1. Analizar casos de uso de Big Data en diversas industrias                      2.2. Relevancia y el impacto de Big Data en la toma de decisiones empresariales.</p>
	3	<p>3. Proyecto de curso                      3.1. Realizar perfil de proyecto                      3.2. Realizar cronograma de proyecto en 12 semanas</p>
<p>Utilizar Tecnologías y Herramientas de Big Data para grandes volúmenes de datos.</p>	4	<p>4. Introducción a Hadoop                      4.1. Conceptos básicos de Hadoop                      4.2. Implementación de la máquina virtual y ambiente Hadoop</p>
	5	<p>4.3. Ejecutar un programa para conteo de palabras con MapReduce</p>
	6	<p>4.4. Localizar y usar los archivos de configuración de HDFS                      4.5. Usar comandos básicos de HDFS                      4.6. Reconstruir datos desde los Snapshots</p>
	7	<p>4.7. Funcionamiento de Hive y su relación con el HDFS                      4.8. Realizar consultas básicas con HiveQL                      4.9. Funcionamiento de las tablas internas y externas</p>
	8	<p>3.3. Presentación de avance de proyecto individual                      3.4. Retroalimentación por pares</p>
	9	<p>4.10. Importación de datos con Sqoop                      4.11. Exportación de datos con Sqoop</p>
	10	<p>4.12. Instalación y configuración de Zookeeper                      4.13. Utilización de Zookeeper</p>
<p>Aplicar Técnicas de Big Data en grandes volúmenes de datos.</p>	11	<p>5. Técnicas de minería de datos                      5.1. Extracción, Transformación y Carga de datos con PySpark                      5.2. Análisis exploratorios de grandes conjuntos de datos para identificar patrones y tendencias</p>
	12	<p>3.5. Presentación de avance de proyecto individual con análisis de datos                      3.6. Retroalimentación por pares</p>
	13	<p>5.3. Instalación de Prerrequisitos: Pytho, Java, Jupyter Notebook.                      5.4. Instalación y puesta en funcionamiento de Apache Spark en Computador propio</p>
	14	<p>6. Minería de datos                      6.1. Técnicas de análisis de datos: minería de datos y aprendizaje automático                      6.2. Análisis exploratorios de grandes conjuntos de datos para identificar patrones y tendencias</p>
	15	<p>7. Análisis predictivo                      7.1. Fundamentos del Análisis predictivo y técnicas de modelado avanzado                      7.2. Modelos de aprendizaje automático para predicción y clasificación utilizando Python o R</p>
	16	<p>3.7. Presentación de proyectos de análisis de Big Data                      3.8. Discusión de la metodología, herramientas, técnicas y resultados obtenidos en los proyectos</p>

## 8. BIBLIOGRAFÍA

### a. BÁSICA / LIBROS

Bibliografía (basarse en normas APA)	Código Biblioteca PUCE	Nro. de ejemplares	Link
Ortega Candel, J. (2023). Big data, machine learning y data science en Python	.312 Or8b, código de barras: P	En línea	<a href="https://puce.odilo.us/info/big-data-machine-learning-y-data-science-en-python-03342899">https://puce.odilo.us/info/big-data-machine-learning-y-data-science-en-python-03342899</a>
Tutorial de Pyspark: Primeros pasos con Pyspark   DataCamp		En línea	<a href="https://www.datacamp.com/es/tutorial/pyspark-tutorial-getting-started-with-pyspark?utm_source=google&amp;utm_medium=paid_search&amp;utm_campaignid=20616617505&amp;utm_adgroupid=154290358037&amp;utm_device=c&amp;utm_keyword=&amp;utm_matchtype=&amp;utm_network=s&amp;utm_adposition=&amp;utm_creative=678190359267&amp;utm_targetid=dsa-2220216603507&amp;utm_loc_interest_ms=&amp;utm_loc_physical_ms=9069516&amp;utm_content=&amp;utm_campaign=231025_1-sea~dsa~tofu_2-b2c_3-es-lang_4-prw_5-na_6-na_7-le_8-gdsh-go_9-nb-s_10-na_11-na&amp;gad_source=2&amp;gclid=EA1aIQobChMk4H4zIqdhwMVJJpaBR2KtwNGEAAAYASAAEgIYrPD_BwE">https://www.datacamp.com/es/tutorial/pyspark-tutorial-getting-started-with-pyspark?utm_source=google&amp;utm_medium=paid_search&amp;utm_campaignid=20616617505&amp;utm_adgroupid=154290358037&amp;utm_device=c&amp;utm_keyword=&amp;utm_matchtype=&amp;utm_network=s&amp;utm_adposition=&amp;utm_creative=678190359267&amp;utm_targetid=dsa-2220216603507&amp;utm_loc_interest_ms=&amp;utm_loc_physical_ms=9069516&amp;utm_content=&amp;utm_campaign=231025_1-sea~dsa~tofu_2-b2c_3-es-lang_4-prw_5-na_6-na_7-le_8-gdsh-go_9-nb-s_10-na_11-na&amp;gad_source=2&amp;gclid=EA1aIQobChMk4H4zIqdhwMVJJpaBR2KtwNGEAAAYASAAEgIYrPD_BwE</a>
		En línea	

### b. COMPLEMENTARIA / CAPÍTULOS EN LIBROS

Bibliografía (basarse en normas APA)	Código Biblioteca PUCESI	Nro. de ejemplares	Link
Minelli, M., & Chambers, M. (2020). Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses. John Wiley & Sons			
Marr, B. (2020). Big Data in Practice: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results. Wiley			
Karau, H., & Warren, R. (2021). High Performance Spark: Best Practices for Scaling and Optimizing Apache Spark. O'Reilly Media			
Aggarwal, C. C. (2021). Data Mining: The Textbook (2nd ed.). Springer			
Vayghan, J. A. (2021). Big Data Governance: Modern Data Management Principles for Hadoop, NoSQL & Big Data Analytics. Independently published			
Fang, X., & Hu, Y. (2021). Big Data applications in China's health sector: Case studies and implications. <i>Data Science and Management</i> , 2, 77-85.			
Kumar, A., & Singh, K. (2022). Big Data analytics in supply chain management: A case study approach. <i>International Journal of Production Research</i> , 60(4), 1230-1245.			
Gupta, S., & Bose, I. (2021). Big Data in finance: Evidence from Asia-Pacific case studies. <i>Journal of Financial Data Science</i> , 3(2), 45-55.			

