Instituto Politécnico Nacional

Secretaría Académica
Dirección de Educación Virtual

Secretaría de Investigación y Posgrado
Dirección de Posgrado

SIP-30

Formato para registro de Unidades de aprendizaje 2021

I.- Datos de identificación de la unidad de aprendizaje

Unidad académica:	Multisede (CIC, CIDETEC, ESCOM, ESFM, UPIITA)									
Programa académico:	Maestría en Ciencia y Tecnología de Inteligencia Artificial y Ciencia de Datos									
		Doctorado						Orientación profesional		
	X	Maestría					X	Orientado a la investigación		
		Especialidad						Con la industria		
								Especialidad médica		
	Sesión de colegio donde se propuso:							Fecha de propuesta:		
Nombre de unidad de aprendizaje:	**Big Data**									
	Clave de la unidad de aprendizaje:						Créditos:	5	*REP 2017*	
	Semanas del semestre	18		Horas a la semana:	4			Horas totales:	72	
Tipo de unidad de aprendizaje:	Obligatoria:	X		Optativa:			Observaciones:			
	Semestre:	1 - 3								
	Teórica (%):	50		Práctica (%)	50			Teórico-prácticas (%):		
Área del conocimiento:	Ingeniería y Ciencias Fisicomatemáticas	X		Ciencias Sociales y Administrativas			Ciencias Medico Biológicas		Interdisciplinario	
Modalidad no escolarizada:	No escolarizada			Nombre de la Plataforma:						
	Mixta			Presencial (%):				En plataforma (%):		
Horas establecidas en el programa de estudios:	Presenciales (si procede) (horas x semana)							En plataforma (horas x semana):		

Formato para registro de Unidades de aprendizaje 2021

I. Aprendizajes que el estudiante deberá demostrar al finalizar

Conocimientos	Habilidades y destrezas	Actitudes y valores
• Características de los problemas de escala y las implicaciones de Big Data en los requisitos de computación • Cuestiones teóricas y metodológicas empleadas en el contexto de Big Data • Algoritmos adecuados para aprovechar la potencia de procesamiento del clúster • Enfoques para simplificar la interfaz de programación utilizada en el desarrollo de aplicaciones Big Data	• Describir las principales líneas de conocimiento necesarias para abordar las aplicaciones de Big Data, destacando las áreas en las que es deseable la colaboración • Proporcionar familiaridad con una variedad de habilidades que pueden usarse en la implementación de aplicaciones de Big Data • Infundir confianza en el manejo de los problemas de Big Data	• Observación crítica a los problemas de big data para seleccionar las técnicas, herramientas y soluciones al problema de Big Data en turno a resolver. • Obtendrá mayor poder de abstracción. • Interacción con personas expertas en los objetos bajo estudio y sus cualidades, que dominan áreas del saber distintas de la informática relacionadas a Big Data. • Ética con respecto al uso de los conjuntos de datos que le confía el cliente o dueño de los datos

Resolución que aborda la propuesta con su enfoque disciplinar

El término 'Big Data' se ha acuñado para describir sistemas que son realmente grandes; estos pueden incluir, por ejemplo, archivos de videos, imágenes, escritura a mano, etc. que no se pueden acomodar en un solo servidor. Dichos sistemas presentan problemas de escala: cómo almacenar grandes cantidades de datos, cómo estar seguros de que los datos son de alta calidad, cómo procesarlos de manera eficiente y cómo obtener información que resulte útil. Estos asuntos se abordan a continuación bajo los títulos de problemas de escala, almacenamiento de datos, computación de alto rendimiento y teoría de la complejidad. Estos temas incluyen una gama de técnicas típicamente utilizadas para abordar los problemas de escala. Dichos sistemas pueden ser complejos, por lo que también se considera el soporte de software para aplicaciones de Big Data.

Formato para registro de Unidades de aprendizaje 2021

II. Proximidad formativa

Áreas multi, inter y transdisciplinarias	Líneas de Generación y Aplicación de Conocimiento	Sectores sociales
• Ciencias de la Computación. • Representación de la información. • Bases de Datos. • Minería de Datos. • Arquitecturas de Big Data	• Aprendizaje automático • Redes neuronales y aprendizaje profundo • Reconocimiento de patrones • Minería de datos, descubrimiento de conocimiento y analítica avanzada • Robótica inteligente • Minería de texto y procesamiento de lenguaje natural	• Los conocimientos, habilidades y destrezas adquiridos en el curso serán de gran ayuda a los tomadores de decisiones a nivel medio y superior. Inteligencia de negocios, gerentes, supervisores, políticos y gobernantes.

Estrategia de asociación:

El instructor deberá de facilitar al estudiante una gran cantidad de datos, así como el problema o cuestión a resolver.
El estudiante aplicará las técnicas aprendidas en este curso para hallar soluciones y mostrarlas de manera que el tomador de decisiones pueda entenderlas fácilmente al tratamiento de grandes volúmenes de datos, donde se requieren que trabajen más de un equipo tanto para almacenar datos como para procesarlos.
Mientras se obtienen las soluciones a problemas planteados en el curso, el estudiante abordará al instructor exponiendo sus dudas sobre los cambios necesarios a los algoritmos de recolección, de procesamiento y de análisis de problemas resueltos con un solo a equipo a problemas que se deben de resolver con un conjunto de equipos de cómputo..

III Metodología de enseñanza – aprendizaje

Descripción
A. Técnica didáctica usada: En tres pasos o fases: (1) exposición de la teoría, teoremas, definiciones, postulados, hipótesis y demostraciones, señalando las inter-relaciones entre los diversos conceptos definidos y los teoremas y restricciones que los enlazan. Atender dudas de los estudiantes. Exposición presencial con ayuda de láminas de Powerpoint, o en línea con exposición síncrona, pero vía internet. (2) El estudiante aplicará la teoría a través de diversas herramientas informáticas, aplicándolas a bases de datos que él o ella organizará en propuestas de solución a problemas de Big Data, obteniendo resultados que serán evaluados por el profesor. Esas evaluaciones serán la base de su calificación. (3) Breve exposición, con láminas de apoyo, de los resultados obtenidos en (2), haciendo énfasis en la administración, procesamiento, análisis y resultados con una arquitectura de Big Data, B. Uso de la teoría subyacente en los temas del curso, ofreciéndole una base sólida al estudiante. C. Uso de software de libre acceso, para practicar los elementos teóricos expuestos en la teoría. D. Uso de bases de datos susceptibles de estructurarse como problemas de Big data, para aplicarles las herramientas estudiadas en el curso,

Formato para registro de Unidades de aprendizaje 2021

consolidando así los conceptos teóricos vertidos en el curso.
La calificación mínima aprobatoria para la Unidad de aprendizaje será ocho (8). Siendo la escala de notas y valores numéricos la siguiente:
Menor a 8: No aprobado
8: Bueno
9: Muy bueno
10: Excelente

Evidencias como proceso de aprendizaje	Evidencias integradoras (resultados que contribuyen al curriculum)	Ponderación
- Lista de asistencia en las 18 semanas - Lista de dudas registradas en las asesorías solicitadas 1. Ejercicios sobre la teoría expuesta en el inciso A.(1) de la metodología señalada en el cuadro III. 2. Descripciones de las herramientas usadas en el inciso A(2) de la metodología señalada en el cuadro III. 3. Descripción de los datos usados en el inciso A(2) de la metodología señalada en el cuadro III. 4. Descripción de los resultados obtenidos en el inciso A(3) de la metodología señalada en el cuadro IV.	• Los alumnos pueden laborar en grupos pequeños, de hasta tres personas, para obtener un resultado. Sin embargo, las presentaciones y hallazgos son individuales, y serán reportados por cada estudiante como se señala en los puntos 1, 2, 3 y 4 del cuadro "Evidencias como proceso de aprendizaje". • Es posible que el alumno integre a su aprendizaje otras herramientas que él desee, y que trabaje sobre otras bases de datos que él considere. Esto con aprobación del profesor.	La suma de las evidencias 2, 3 y 4 valdrá 80 puntos. Los ejercicios del punto 1 valen 20 p.

IV. Descripción de la participación esperada en el estudiante

Receptiva	Resolutiva	Autónoma	Estratégica

Formato para registro de Unidades de aprendizaje 2021

• Atenderá las exposiciones del inciso A.(1) de la metodología señalada en el cuadro III. • Usará las herramientas del inciso A(2), cuadro III. • Usará las bases de dastos del inciso A(3), cuadro III. • Podrá usar otras, según el cuadro "Evidencias integradoras".	• El alumno entregará respuestas a los ejercicios A(1), cuadro III • El alumno entregará resultados obtenidos en A(2) y A(3), cuadro III (Metodología de enseñanza-aprendizaje) • Un elemento secundario pero de cierta importancia es la claridad en su exposición y la precisión de sus presentaciones.		Los proyectos están abiertos a lo que proponga el alumno.

Contenido temático

1. Problemas de Escala (8 horas)
   a) La necesidad de medición en el contexto de Big Data, incluido el tamaño, la capacidad y el tiempo
   b) El concepto del tamaño de un problema
   c) Consecuencias de la rápida tasa de crecimiento a considerar para el cálculo
   d) Consecuencias de almacenamiento de la rápida tasa de crecimiento de datos
   e) La necesidad de poner énfasis en la simplicidad
   f) Enfoques para abordar los problemas de coordinación con un número creciente de agentes/procesos
   g) Enfoques para abordar los problemas de escala mientras se acomoda la escalabilidad

2. Arquitecturas informáticas de Big Data (8 horas)
   a) Mecanismos que soportan entradas/salidas rápidas y eficientes
   b) Los conceptos y requisitos de la computación de alto rendimiento centrada en datos
   c) Consideraciones de memoria: consideraciones de caché, incluida la coherencia de caché
   d) Las diversas arquitecturas de cómputo paralelo, sus fortalezas y sus limitaciones: multinúcleo, computación grid, GPU, memoria compartida, memoria distribuida, multiprocesamiento simétrico, procesamiento vectorial
   e) Taxonomía de Flynn
   f) Consideraciones de instrucción en apoyo del paralelismo

Formato para registro de Unidades de aprendizaje 2021

g) Jerarquía de almacenamiento en paralelo

3. Frameworks de Computación Paralela (8 horas)
   a) Definición y propósito de un modelo de computación paralela
   b) Clasificación de modelos
   c) Sistemas distribuidos
   d) Búsqueda de cuadrícula
   e) Interacción de procesos: cuestiones de comunicación y coordinación
   f) Descomposición de problemas: descomposición basada en tareas, descomposición en paralelo de datos

4. Almacenamiento de datos distribuidos (8 horas)
   (a) Enfoques para almacenar grandes cantidades de datos, incluido el almacenamiento en una variedad de dispositivos
   (b) Jerarquías de almacenamiento
   (c) Asegurar datos limpios, consistentes y representativos
   (d) Protección y mantenimiento de los datos
   (e) Problemas de recuperación
   (f) Los beneficios y limitaciones de una variedad de técnicas utilizadas para abordar los problemas de escala, como hashing, filtrado, muestreo
   (g) Copias de seguridad

5. Programación Paralela (8 horas)
   (a) Concurrencia, paralelismo y sistemas distribuidos
   (b) Limitaciones del paralelismo incluyendo los gastos generales
   (c) Algoritmos paralelos y cómo se adaptan mejor a arquitecturas de hardware particulares; problemas de equilibrio de carga
   (d) Paradigma típico de programación paralela como MapReduce
   (e) Complejidad de algoritmos paralelos/concurrentes

6. Técnicas para aplicaciones de Big Data (8 horas)
   (a) La necesidad de técnicas para ayudar con el manejo de Big Data
   (b) Hashing, Sampling, filtrado y sus limitaciones

Formato para registro de Unidades de aprendizaje 2021

(c) Bosquejo de datos y sinopsis

7. Computación en la nube (8 horas)
   (a) La naturaleza de la computación en la nube y sus ventajas
   (b) La arquitectura de un centro de datos
   (c) Riesgos asociados con la computación en la nube
   (d) Diferentes enfoques para respaldar la computación en la nube
   (e) Sistemas de archivos distribuidos
   (f) Servicios en la nube en apoyo de aplicaciones de Big Data
   (g) Tecnología de virtualización
   (h) Problemas de seguridad para la nube, incluida la computación en la nube, el almacenamiento en la nube y las máquinas virtuales

8. Teoría de la Complejidad (8 horas)
   (a) Problemas de computación y eficiencia de los algoritmos
   (b) La noción de complejidad computacional, su uso en el contexto de concurrencia/paralelismo y su importancia en el contexto de Big Data
   (c) Limitaciones al concepto de complejidad
   (d) Evaluación de la complejidad de una variedad de algoritmos de uso común, incluidos aquellos que exhiben concurrencia/paralelismo

9. Soporte de software para aplicaciones de Big Data (8 horas)
   (a) La necesidad de entornos de programación para soportar aplicaciones Big Data y la naturaleza de estos
   (b) Conceptos de escalado automático y computación sin servidor
   (c) Revisión de la disponibilidad de servicios web sofisticados para el soporte del movimiento de datos, análisis y aprendizaje automático en el contexto de Big Data

V. Secuencia programática

No.	Tema	Objetivo de aprendizaje / competencia específica	Tiempo/Horas/Semanas

Formato para registro de Unidades de aprendizaje 2021

1				
Actividad(es):			Tipo de interacción(es):	
			Referencias (s):	
Evidencia(s):				

VI. Habilitadores tecnológicos

	Disposiciones	Especificaciones / descripción de efectos
	Conectividad	Conectarse a diferentes fuentes de datos con diferentes formatos
	Habilidades digitales	Uso de bases de datos relacionales con soltura. Manejo de dos lenguajes de programación modernos (Java, R, Python, u otro similar). Manejo o familiarización con conceptos estadísticos.
	Interoperabilidad	Poca interoperabilidad. Los datos que están en plataformas exógenas serán copiados a repositorios digitales locales, propiedad del estudiante o del CIC, lo mismo que las herramientas de uso libre.
	Datos abiertos	No se usarán, salvo que alguna base de datos abiertas pueda para problemas de análisis de Big Data con las técnicas de este curso.
	Big Data	El curso prepara al estudiante para analizar grandes bases de datos, pero la didáctica del curso limita las bases de datos empleadas a ser de tamaño modesto, para aprovechar las técnicas de paralelismo, distribución y procesamiento en conjuntos de equipos de cómputo y almacenamiento coordinados.
	Machine Learning	El curso es parte del programa de aprendizaje mecánico, que se lleva a cabo en Laboratorios de Ciencia de Datos.
	Simulación	
	Realidad aumentada	
	Otro…	

VII. Referencias

Formato para registro de Unidades de aprendizaje 2021

Conferencias magistrales	Notas complementarias
1.	
2.	
3.	

Documentales / electrónicas

4. Jure Leskovec. Anand Rajaraman, Jeffrey D. Ullman, “The Mining of Massive Datasets book”, 3th Edition, 2020
5. Jiawei Han and Micheline Kamber , “Data Mining: Concepts and Techniques” , Third edition The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor , Morgan Kaufmann Publishers, March 2011. ISBN 1-55860-901-6
6. Joyanes, L. *Big Data: Análisis de grandes volúmenes de datos en organizaciones.* México: Alfaomega, 2016.
7. Michael Kaufmann, Andreas Meier , SQL and Nosql Databases: Modeling, Languages, Security and Architectures for Big Data, Springer 2017.
8.

VIII. Créditos y responsivas

	Responsabilidad	Nombre completo	Clave de nombramiento /No. de empleado
	Coordinador (Autor)	Gilberto Lorenzo Martínez Luna	16300-EH-22/9601000
	Participante (Coautor)	Adolfo Guzmán Arenas	13702-EG-18/6/9500093
	Asesor didáctico / Diseñador Instruccional		
	Tecnólogo educativo / Comunicólogo		
	Corrector de estilo		
	Programador multimedia / Diseñador gráfico		
Otro…			

Instituto Politécnico Nacional

Secretaría Académica
Dirección de Educación Virtual

Secretaría de Investigación y Posgrado
Dirección de Posgrado

SIP-30

Formato para registro de Unidades de aprendizaje 2021

VERIFICACIÓN GENERAL DE LA PLANEACIÓN DIDÁCTICA	REVISIÓN DE LA PLANEACIÓN DIDÁCTICA (VIABILIDAD)
Por la División de Operación y Promoción al Posgrado de la SIP Nombre_____ FIRMA_____	Por la Subdirección de Diseño y Desarrollo de la DEV Nombre_____ FIRMA_____

VERIFICACIÓN PARA SU PUESTA EN OPERACIÓN	REVISIÓN TÉCNICO-PEDAGÓGICA PARA LA MODALIDAD
Por la Dirección de Posgrado Nombre_____ FIRMA_____ SELLO DE VALIDACIÓN	Por la Dirección para la Educación Virtual Nombre_____ FIRMA_____