

# Programa de Asignatura

## 1. Identificación Asignatura

<b>Nombre:</b>	Big Data		<b>Código:</b>	IF1026-1
<b>Carrera:</b>	Ingeniería Civil Informática	<b>Unidad Académica:</b>	Ciencias Naturales y Tecnología	
<b>Ciclo Formativo:</b>	Ciclo Licenciatura	<b>Línea formativa:</b>	Especializada	
<b>Semestre</b>	VIII	<b>Tipo de actividad:</b>	Obligatoria	
<b>N° SCT:</b>	6	<b>Horas Cronológicas Semanales</b>		
		<b>Presenciales:</b>	4.5	<b>Trabajo Autónomo:</b>
<b>Pre-requisitos</b>	Visualización de Datos			

## 2. Propósito formativo

El curso de Big Data tiene como propósito que los estudiantes adquieran los conceptos fundamentales y avanzados necesarios para gestionar grandes cantidades de datos, tanto estructurados como no estructurados. Durante el curso, los estudiantes desarrollarán una comprensión profunda de las tecnologías y técnicas críticas en el campo del Big Data, y aplicarán estos conocimientos en el desarrollo de soluciones a problemas complejos relacionados con la captura, gestión y visualización de grandes volúmenes de datos.

Los objetivos del curso incluyen la adquisición de conocimientos fundamentales sobre los principios y conceptos esenciales del Big Data, como las características de los datos, los paradigmas de procesamiento y las tecnologías subyacentes. Además, los estudiantes se familiarizarán con las herramientas y plataformas más utilizadas en Big Data, tales como Hadoop, Spark y bases de datos NoSQL, y aprenderán a implementarlas en proyectos prácticos.

El curso también se enfoca en el desarrollo de competencias prácticas para diseñar, desarrollar e implementar soluciones de Big Data que aborden problemas reales en diversas industrias. Asimismo, se enfatiza la analítica de datos y la visualización, utilizando técnicas de minería de datos y aprendizaje automático para extraer información valiosa de grandes conjuntos de datos.

Finalmente, el curso aborda los desafíos éticos y de privacidad asociados con el manejo de grandes volúmenes de datos, enseñando a los estudiantes a gestionarlos de manera responsable. Esta estructura integral asegura que los estudiantes no solo adquieran conocimientos teóricos, sino también habilidades prácticas y aplicables en el mercado laboral actual.

Esta asignatura es el último curso de la línea de formación en Ciencia de Datos. Se encuentra en el VIII semestre, previo al inicio de cursos en los que estos conocimientos pueden ser importantes, como Gestión Estratégica de la Información y Taller de Título.

## 3. Contribución al perfil de egreso

Esta asignatura contribuye a los siguientes desempeños o resultados de aprendizaje globales declarados en el Perfil de Egreso de la carrera:

1. Gestiona datos de distintas características, asegurando su calidad, acceso eficiente y seguridad, a través de mecanismos que habiliten su almacenamiento y que sean apropiados a las responsabilidades establecidas para su uso.
2. Genera información relevante, por medio de métodos analíticos y experimentales, a partir de fuentes de datos de gran volumen y diversa complejidad, utilizando técnicas pertinentes de minería de datos, con el fin de apoyar la toma de decisiones.

Esta estructura del curso está diseñada para asegurar que los estudiantes no solo adquieran conocimientos teóricos, sino también habilidades prácticas y aplicables en el mercado laboral actual. Se hace hincapié en la integración de tecnologías emergentes y metodologías avanzadas para preparar a los estudiantes para los desafíos del mundo real en el campo del Big Data.

#### 4. Resultados de aprendizaje específicos

Resultado de Aprendizaje Específico	Criterios de evaluación	Evidencia
<b>RA1.</b> Conocer los conceptos fundamentales asociados a Big Data	1.1. Describe características y etapas del procesamiento de datos en Big Data. 1.2. Reconoce las diversas áreas de aplicación del Big Data.	Laboratorios, guías de ejercicio.
<b>RA2.</b> Aprende a través de diferentes tecnologías Big Data, a modelar, recolectar, almacenar y organizar los datos en una solución de alto volumen de datos.	2.2. Modela adecuadamente una solución que requiere gestión de altos volúmenes de datos. 2.2 Implementa una solución a un problema de Big Data sobre una base de datos NoSQL.	Laboratorios, guías de ejercicio.
<b>RA3.</b> Conoce diversas estrategias para implementar analítica en Big Data.	3.1. Aplica ejercicios simples de predicción utilizando Machine learning 3.2. Implementa modelos de analítica sobre alguna problemática de Big Data.	

#### 5. Unidades de Aprendizaje

<p><b>1. Introducción y fundamentos al Big Data</b></p> <p>1.1. Historia y conceptos y características de Big Data 1.2. Los 5 V's de Big Data: Volumen, Velocidad, Variedad, Veracidad, y Valor. 1.3. Paradigmas de procesamiento de datos 1.4. Dimensiones de Big Data 1.5. Arquitecturas, Modelos y Aplicaciones de Big Data</p>
<p><b>2. Tecnologías En Big Data</b></p> <p>2.1. Bases de datos NoSQL 2.2. Distribución e integración de datos 2.3. Herramientas de adquisición de datos (kafka, flume, sqoop, etc) 2.4. Procesamiento Batch o procesamiento por lotes 2.5. Procesamiento en Tiempo Real o en Streaming</p>
<p><b>3. Analítica en Big Data</b></p> <p>3.1. ETL (Extract, Transform, Load): Conceptos y herramientas 3.2. Almacenamiento y consulta de datos 3.3. Analítica de textos, aplicación de ML 3.4. Analítica Web, aplicación de ML 3.5. Analítica avanzada, aplicación de ML 3.6. Ética y privacidad en el manejo de Big Data</p>

## 6. Recursos de Aprendizaje

### Bibliografía:

**B1:** Albert Y. Zomaya Sherif Sakr, “Handbook of Big Data Technologies”, 2017

**B2:** EMC Education Services, “Data Science and Big Data Analytics”, 2015

[https://aitskadapa.ac.in/e-](https://aitskadapa.ac.in/e-books/AI&DS/BIG%20DATA/Data%20Science%20%20Big%20Data%20Analytics%20(%20PDFDrive%20).pdf)

[books/AI&DS/BIG%20DATA/Data%20Science%20 %20Big%20Data%20Analytics%20\(%20PDFDrive%20\).pdf](https://aitskadapa.ac.in/e-books/AI&DS/BIG%20DATA/Data%20Science%20%20Big%20Data%20Analytics%20(%20PDFDrive%20).pdf)

**B3:** Dirk P. Kroese, Zdravko I. Botev, Thomas Taimre, Radislav Vaisman, “Data Science and Machine Learning”, 2023

<https://people.smp.uq.edu.au/DirkKroese/DSML/DSML.pdf>

### Recursos materiales e infraestructura:

- Laboratorio de computación.
- Acceso a Ucampus.
- Acceso a Googlesites con credenciales institucionales.

Computadores debidamente equipados para utilizar lenguajes de alto nivel (por ej.: Python).

## 7. Comportamiento y ética académica:

Se espera que, en el marco de sus actividades académicas y estudiantiles, los estudiantes demuestren un compromiso inquebrantable con los estándares de ética y honestidad académica que distinguen a nuestra comunidad universitaria. Este compromiso debe reflejarse en todas las acciones y trabajos realizados dentro del ámbito de la Universidad de Aysén, siguiendo rigurosamente las directrices establecidas en el Reglamento de Estudiantes, particularmente lo indicado en los artículos 23°, 24°, y 26°.

Además, es imperativo que todas las acciones se alineen con los reglamentos institucionales aplicables y las normativas internas vigentes, incluido el código de ética de la universidad. Estos documentos no solo rigen nuestros comportamientos cotidianos sino que también establecen las expectativas para nuestra convivencia y trabajo académico.

Cualquier infracción a estos principios de honestidad académica, detectada durante la realización, presentación, o entrega de cualquier actividad evaluativa del curso, acarreará consecuencias serias. Estas incluyen la suspensión inmediata de la actividad en cuestión y la asignación de la nota mínima posible (1.0) como reflejo de la gravedad del acto. Esta política se aplica de manera estricta y sin excepciones, reafirmando nuestro compromiso con una formación académica íntegra y responsable.

## Planificación del curso



### 8. Responsables:

<b>Académico (s) Responsable (s) y equipo docente</b>	<b>Profesor:</b> Johnny Valencia C		
<b>Contacto</b>	<b>Correo:</b> <a href="mailto:johnny.valencia@uaysen.cl">johnny.valencia@uaysen.cl</a>		
<b>Año</b>	2024	<b>Periodo Académico</b>	Segundo semestre
<b>Horario clases</b>	<b>Cátedra:</b> Lu 16.15 – 17.45 hrs. Mi 14.30 – 17.45 hrs.  <b>Sesión Prácticas:</b> Lu 16.15 – 17.45 hrs.	<b>Horario de atención estudiantes</b>	Contactar previamente al profesor mediante email.  Disponible Ma 14.30 – 16.00 Vi 14.30 – 16.00
<b>Sala / Campus</b>	Sala Virtual Ucampus		

### 9. Metodología de Trabajo:

En el curso se contemplan cuatro tipos de actividades docentes, las cuales se asocian a requerimientos de sala y al nivel de intervención del profesor/ayudante:

Actividad docente	Descripción	Intervención del profesor/ayudante	Requerimiento de sala
Exposición conceptual	El profesor introduce los fundamentos preliminares de Big Data y conceptos esenciales para el desarrollo de habilidades prácticas, con una participación activa en el aula tradicional o entornos virtuales.	Alta	Sala de clases UCampus Online UCampus Offline
Programación expositiva	Mediante la resolución de problemas específicos de Big Data, el profesor guía a los estudiantes a profundizar en el entendimiento de conceptos clave, empleando ejemplos prácticos en un entorno de sala de clases o plataformas online.	Alta	Sala de clases UCampus Online UCampus Offline
Programación tutorial	Esta modalidad incorpora pausas estratégicas en la exposición, permitiendo a los estudiantes completar tareas específicas bajo la supervisión del docente, favoreciendo el aprendizaje interactivo en laboratorios de computación o con uso de equipos personales	Media	Laboratorio de computación Computador persona
Actividad práctica / Programación autónoma	Fomenta la independencia de los estudiantes al enfrentar y resolver desafíos de Big Data, en grupos o individualmente, con mínima intervención docente, ideal para el trabajo en laboratorios de computación o mediante dispositivos personales.	Baja	Laboratorio de computación Computador persona

Durante el semestre, **se programarán evaluaciones menores** en cualquier momento, con el fin de revisar y reforzar los conocimientos adquiridos hasta la fecha. Estas evaluaciones podrán adoptar diversas formas:

- **Control (Quiz):** Evaluación, ya sea individual o en grupo, realizada en papel durante los primeros 45 minutos de la clase.
- **Trabajo en Clase:** Actividad evaluativa individual o grupal, llevada a cabo en computadora. Podrá ocupar los primeros 45 minutos o extenderse durante toda la sesión.

- **Entrega de Tareas:** Evaluación grupal realizada en computadora y fuera del horario de clase, promoviendo el trabajo colaborativo y la gestión del tiempo.

Adicionalmente, como componente de la calificación final, los estudiantes deberán presentar, un proyecto que abarque el análisis, diseño e implementación de una solución tecnológica. Para esta tarea, se proporcionará un instructivo detallado que guiará a los estudiantes en su preparación y entrega, el desarrollo del proyecto se gestará durante todo el semestre, evidenciando una metodología clara de forma progresiva.

## 10. Evaluaciones:

Evaluación	Ponderaciones específicas	Ponderación nota presentación
Pruebas de cátedra	<ul style="list-style-type: none"> <li>• Prueba 1 (P1): 50%</li> <li>• Prueba 2 (P2): 40%</li> <li>• Quiz (Q1): 10%</li> </ul>	60%
Evaluaciones menores	El promedio simple entre las notas consideradas corresponde al 100%.	40%

### Calificación final:

- Nota de presentación: 70%
- Examen Final: 30 % (Exposición de proyecto de aplicación)

### Requisito de Aprobación

- Asistencia: 65%
- Nota Final: 4,0
- La entrega de cualquiera de las evaluaciones indicadas en el programa de la asignatura por fuera del plazo definido serán evaluadas con nota mínima (1,0).
- Sólo para el caso de las evaluaciones presenciales, se permitirá comenzar con un atraso máximo de 15 minutos después de comenzada la misma. El tiempo de retraso no se podrá recuperar. Aquel/la estudiante que se presente a rendir la evaluación después de pasados los 15 minutos, se evaluará con nota mínima (1,0).
- En caso de inasistencia o no entrega en plazo de alguna evaluación, se podrán justificar mediante el mismo procedimiento antes descrito para la justificación de inasistencias actividades lectivas.
- Las clases teóricas tienen un 65% de asistencia mínima obligatoria, siguiendo el mínimo requerido por el Reglamento General de Estudios de Pregrado. La asistencia a las clases prácticas (laboratorios y terrenos) es del 100%. El no cumplimiento de estos porcentajes de asistencia será causal de reprobación de la asignatura.
- En casos debidamente justificados ante el Registro Académico, el/ la estudiante que no haya asistido a una salida a terreno o laboratorio tendrá derecho a rendir examen. Se considerarán debidamente justificadas las inasistencias ante el Registro Académico aquéllas que estén respaldadas con certificados médicos, laborales o algún documento validado por la Unidad de Acceso y Desarrollo Estudiantil. Las actividades de terrenos y laboratorios no podrán ser recuperadas.
- Se permitirá el ingreso posterior a la hora de inicio con un máximo de 15 minutos, siempre y cuando no sea una acción repetida por la/el estudiante (se aceptará máximo de 3 veces). Para las salidas a terreno, el tiempo de espera máximo será de 5 minutos. Para casos donde las actividades lectivas contemplen más de un módulo, el/la estudiante que no haya asistido al módulo anterior, podrá ingresar al comienzo del nuevo módulo.
- En caso de que ningún estudiante se presente a la actividad lectiva después de 15 minutos de comenzada, ésta se suspenderá. Los contenidos programados para dicha actividad se darán por dictados, será responsabilidad del estudiante ponerse al día con los contenidos de dicha clase. Los contenidos de dicha clase, y ejercicios, si así lo hubiera, serán enviados para ser realizados como trabajo autónomo.

### Otros aspectos asociados al funcionamiento del curso:

- Toda la coordinación del curso (comunicaciones, actualización de notas, material, etc.) será realizada a través de UCampus.
- **Es deber del estudiante mantenerse informado de las noticias, avisos y material entregado por el profesor a través de estos medios, se sugiere instalar en su dispositivo móvil la aplicación de UCampus.**
- Recordar que los correos electrónicos serán respondidos en horario laboral (lunes a viernes de 9:00 a 18:00hs), no se responderán correos fuera de ese horario.
- Durante el desarrollo de las actividades lectivas, los teléfonos celulares deberán estar en silencio y guardados, a menos que el/la profesor/a específicamente requiera de estos equipos para la realización de su clase o durante algunos casos excepcionales conversados previamente con el/la docente a cargo.
- Las actividades lectivas se dictarán de forma presencial, salvo excepciones sujeto a contingencias presentes durante el transcurso de la asignatura.

### 11. Planificación de las actividades de enseñanza- aprendizaje y de evaluación:

Semana	Resultado (s) de Aprendizaje (*)	Tema (Unidades de aprendizaje) y actividades	Recursos utilizados o lecturas (*)	Actividad(es)
1	RA1	<b>Introducción y Fundamentos al Big Data</b> 1.1. Historia, conceptos y características de Big Data 1.2. Los 5 V's de Big Data: Volumen, Velocidad, Variedad, Veracidad, y Valor	B1, B2, B3	
2	RA1	<b>Introducción y Fundamentos al Big Data</b> 1.1. Historia, conceptos y características de Big Data 1.2. Los 5 V's de Big Data: Volumen, Velocidad, Variedad, Veracidad, y Valor	B1, B2, B3	
3	RA1	<b>Paradigmas y Dimensiones de Big Data</b> 1.3. Paradigmas de procesamiento de datos 1.4. Dimensiones de Big Data	B1, B2, B3	
4	RA1, RA2	<b>Paradigmas y Dimensiones de Big Data</b> 1.3. Paradigmas de procesamiento de datos 1.4. Dimensiones de Big Data	B1, B2, B3	
5	RA1, RA2	<b>Arquitecturas y Modelos de Big Data</b> 1.5. Arquitecturas, Modelos y Aplicaciones de Big Data	B1, B2, B3	
6	RA1, RA2	<b>Arquitecturas y Modelos de Big Data</b> 1.5. Arquitecturas, Modelos y Aplicaciones de Big Data	B1, B2, B3	Prueba (P1) (9 de septiembre del 2024)
7	RA1, RA2	<b>Tecnologías en Big Data - Bases de Datos NoSQL</b> 2.1. Bases de datos NoSQL	B1, B2, B3	

8	RA1, RA2	<b>Distribución e Integración de Datos</b> 2.2. Distribución e integración de datos	B1, B2, B3	
9	RA2, RA3	<b>Herramientas de Adquisición de Datos</b> 2.3. Herramientas de adquisición de datos (Kafka, Flume, Sqoop, etc)	B1, B2, B3	
10	RA2, RA3	<b>Herramientas de Adquisición de Datos</b> 2.3. Herramientas de adquisición de datos (Kafka, Flume, Sqoop, etc)	B1, B2, B3	
11	RA2, RA3	<b>Procesamiento Batch o por Lotes</b> 2.4. Procesamiento Batch o procesamiento por lotes	B1, B2, B3	Prueba (P2) (23 de Octubre del 2024)
12	RA2, RA3	<b>Procesamiento en Tiempo Real o en Streaming</b> 2.5. Procesamiento en Tiempo Real o en Streaming		
13	RA2, RA3	<b>ETL y Almacenamiento de Datos</b> 3.1. ETL (Extract, Transform, Load): Conceptos y herramientas 3.2. Almacenamiento y consulta de datos	B1, B2, B3	
14	RA2, RA3	<b>Analítica de Textos y Web con ML</b> 3.3. Analítica de textos, aplicación de ML 3.4. Analítica Web, aplicación de ML	B1, B2, B3	
15	RA3	<b>Analítica Avanzada y Ética en Big Data</b> 3.5. Analítica avanzada, aplicación de ML 3.6. Ética y privacidad en el manejo de Big Data	B1, B2, B3	Quiz 1 (Q1) (13 de noviembre del 2024)
16	RA3	<b>Revisión y Proyectos Finales</b> <ul style="list-style-type: none"> <li>• Repaso de conceptos clave</li> <li>• Presentación de proyectos finales</li> </ul>	B1, B2, B3	
17		Periodo de exámenes		Exposición del proyecto de aplicación (5 al 18 de Diciembre del 2024)
18		Periodo de exámenes		Exposición del proyecto de aplicación (5 al 18 de Diciembre del 2024)

(\*) Para referencias de Resultados de Aprendizaje y Recursos, consultar el programa del curso.