

## Sílabo

### 170216 - Big Data Analytics

#### I. Información general

Nombre del Curso: Big Data Analytics  
Código del curso: 170216  
Departamento Académico: Ingeniería  
Créditos: 4  
Horas Teoría: 4  
Horas Práctica: 0  
Periodo Académico: 2023-01-PRE  
Sección: A  
Modalidad: Presencial  
Idioma: Español  
Docente: JUNIOR JOHN FABIAN ARTEAGA  
Email docente: jj.fabiana@up.edu.pe

#### II. Introducción

En el presente curso se estudian como desarrollar soluciones mediante el uso de tecnología que implica conocer cómo elaborar los requerimientos, casos de uso, arquitectura de una solución Big Data alineada con la estrategia de la empresa para optimizar un proceso determinado al seno de un empresa.

#### III. Logro de aprendizaje final del curso

Al término del curso, el estudiante sustentará un informe en el que presenta los requerimientos, casos de uso, arquitectura de una solución Big Data alineada a la propuesta de la estrategia de una empresa y el caso de negocio. Demostrará también que la arquitectura planteada es coherente con el caso de negocio.

#### IV. Unidades de aprendizaje

##### Unidad de aprendizaje 1: Fundamentos de Big Data

###### Logro de Aprendizaje / propósito de la unidad:

Comprende los conceptos generales sobre Big Data

###### Contenidos:

- Conceptos básicos de Big Data.
- Conceptos básicos de Apache Hadoop

##### Unidad de aprendizaje 2: Plataformas de Big Data I

###### Logro de Aprendizaje / propósito de la unidad:

Conoce y aplica las principales tecnologías existentes en el manejo y análisis de Big Data

###### Contenidos:

- MapReduce.
- Pig.
- Hive.

### Unidad de aprendizaje 3: Plataformas de Big Data II

**Logro de Aprendizaje / propósito de la unidad:**

Desarrolla la capacidad de poder crear estrategias de analítica de Big Data para casos reales a partir de los conceptos desarrollados en el curso y/o relacionados.

**Contenidos:**

- HBase.
- Spark.

### Unidad de aprendizaje 4: Técnicas de Machine Learning en Big Data

**Logro de Aprendizaje / propósito de la unidad:**

Conoce y aplica las principales técnicas de minería de datos y aprendizaje de máquinas en el análisis de Big Data

**Contenidos:**

- Machine Learning en Big Data.
- PySpark.

## V. Estrategias Didácticas

Cuestionarios y asignaciones basados en la materia de clase y lecturas asignadas, los cuales son evaluados.

Participación activa de parte de los estudiantes mediante la solución compartida de problemas.

Las clases prácticas se llevarán a mediante la discusión y resolución de casos.

## VI. Sistemas de evaluación

**Consideraciones para las evaluaciones**

El Promedio Final (PF) del curso se obtiene de la siguiente forma:

$$PF = 0.25*EP01 + 0.35*EF02 + 0.40*NT03$$

- EP01 es la nota del Examen Parcial
- EF02 es la nota del Examen Final
- NT03 es la nota de Trabajos.

La Nota de Trabajo se calcula de la siguiente forma:

$$NT03 = (0.40*PC01 + 0.40*PC02 + 0.2*Otras_Actividades)*Factor\_Proyecto$$

- PC01 y PC02 son las notas de las Prácticas Calificadas, respectivamente.
- Otras\_Actividades: representa el promedio de las notas correspondientes al desarrollo de tareas, controles de lectura, participación, laboratorios y similares.
- Factor\_Proyecto: refleja la contribución del proyecto del curso y se calcula según los valores de equivalencia de la Tabla 1 con la Nota\_Proyecto\_Integrador.

Nota\_Proyecto: es la nota del proyecto integrador del curso. Se calcula de acuerdo con la siguiente formula:



$Nota\_Proyecto = (0.4*Entrega\_1 * 0.6*Entrega\_2 + Modificador\_Exposición)*Factor\_Evaluación\_Pares$

- Entrega\_1 y Entrega\_2 son las notas correspondientes al avance y presentación final del proyecto.
- Modificador\_Exposición: acorde a la sustentación individual, puede adquirir los siguientes valores:
  - Sobresaliente: +1
  - Suficiente : 0
  - Insuficiente : Desde -1 hasta -10 según el criterio del docente.
- Factor\_Evaluación\_Pares: refleja la nota de contribución (en cantidad y calidad) dada por el resto de los integrantes del grupo. Esta nota puede adquirir los siguientes valores:
  - Aporte sobresaliente : 1.05
  - Aporte suficiente : 1.00
  - Aporte regular : 0.90
  - Aporte insuficiente : 0.80

- Valores de equivalencia

Nota Vigesima	Factor Equivalente
0	0.30
01 - 02	0.50
03 - 05	0.60
06 - 08	0.70
09 - 11	0.80
12 - 14	0.90
15 - 17	1.00
18 - 20	1.05

Tabla 1

- AVISOS IMPORTANTES

- La ausencia injustificada a la exposición del Proyecto convierte la Nota\_Proyecto en cero.
- No hay puntos extras en Prácticas Calificadas ni en los Exámenes Parcial o Final bajo ningún motivo. Cualquier punto extra será considerado en Otras\_Actividades.
- Considerando el nuevo Reglamento de Estudiantes, en este curso no se aplicará suspensión por inasistencia.
- En el siguiente enlace se encuentra disponible el sistema de evaluación y las disposiciones para la evaluación de rezagados (<https://sites.google.com/alum.up.edu.pe/sistemadeevaluacionestandar01/inicio>)

Nombre evaluación	%	Fecha	Criterios	Comentarios
1. Examen Parcial	25		Conoce y aplica los conceptos aprendidos durante la primera parte del curso	
2. Examen Final	35		Conoce y aplica los conceptos aprendidos durante la segunda parte del curso	
3. Nota Trabajos	40		Incluye: 2 Prácticas calificadas, Proyecto Final y Otras actividades (casos, actividades individuales/grupales)	

## VII. Cronograma referencial de actividades

Unidades de aprendizaje	Contenidos y actividades a realizar	Recursos y materiales	Evaluaciones
<b>Semana 1: del 20/03/2023 al 25/03/2023</b>			
• Unidad de aprendizaje 1: Fundamentos de Big Data	• Presentación del curso. • Big Data: Preliminares. • Aplicaciones.		
<b>Semana 2: del 27/03/2023 al 01/04/2023</b>			
• Unidad de aprendizaje 1: Fundamentos de Big Data	• Dimensiones de Big Data • Áreas relacionadas. • Programación.	• Lecturas	
<b>Semana 3 con feriados el jueves 06, viernes 07 y sábado 08: del 03/04/2023 al 08/04/2023</b>			
• Unidad de aprendizaje 1: Fundamentos de Big Data	• Sistema de Archivos. • DFS. • HDFS.  <b>Propuesta de Proyecto</b>		
<b>Semana 4: del 10/04/2023 al 15/04/2023</b>			
• Unidad de aprendizaje 1: Fundamentos de Big Data • Unidad de aprendizaje 2: Plataformas de Big Data I	• Apache Hadoop. • YARN. • Map-Reduce.	• Casos	
<b>Semana 5: del 17/04/2023 al 22/04/2023</b>			
• Unidad de aprendizaje 2: Plataformas de Big Data I	• Cloudera. • Comandos básicos Linux. • Comandos HDFS en CentOS.		
<b>Semana 6: del 24/04/2023 al 29/04/2023</b>			
• Unidad de aprendizaje 2: Plataformas de Big Data I	• Tecnologías de Big Data. • Apache Pig. <b>PC1</b>		• Nota Trabajos
<b>Semana 7: del 01/05/2023 al 06/05/2023</b>			

Unidades de aprendizaje	Contenidos y actividades a realizar	Recursos y materiales	Evaluaciones
<ul style="list-style-type: none"> <li>Unidad de aprendizaje 2: Plataformas de Big Data I</li> </ul>	<ul style="list-style-type: none"> <li>Tecnologías de Big Data.</li> <li>Apache Hive.</li> </ul> <b>Avance de Proyecto</b>		<ul style="list-style-type: none"> <li>Nota Trabajos</li> </ul>
<b>Semana 8 de exámenes parciales: del 08/05/2023 al 13/05/2023</b>			
<ul style="list-style-type: none"> <li>Unidad de aprendizaje 1: Fundamentos de Big Data</li> <li>Unidad de aprendizaje 2: Plataformas de Big Data I</li> </ul>	<b>Examen Parcial</b>		<ul style="list-style-type: none"> <li>Examen Parcial</li> </ul>
<b>Semana 9: del 15/05/2023 al 20/05/2023</b>			
<ul style="list-style-type: none"> <li>Unidad de aprendizaje 3: Plataformas de Big Data II</li> </ul>	<ul style="list-style-type: none"> <li>Tecnologías de Big Data.</li> <li>Apache HBase.</li> </ul>		
<b>Semana 10: del 22/05/2023 al 27/05/2023</b>			
<ul style="list-style-type: none"> <li>Unidad de aprendizaje 3: Plataformas de Big Data II</li> </ul>	<ul style="list-style-type: none"> <li>Tecnologías de Big Data.</li> <li>Spark.</li> </ul>		
<b>Semana 11: del 29/05/2023 al 03/06/2023</b>			
<ul style="list-style-type: none"> <li>Unidad de aprendizaje 3: Plataformas de Big Data II</li> <li>Unidad de aprendizaje 4: Técnicas de Machine Learning en Big Data</li> </ul>	<ul style="list-style-type: none"> <li>Librerías para procesamiento numérico</li> <li>PySpark para importación/tratamiento de datos.</li> </ul>		
<b>Semana 12: del 05/06/2023 al 10/06/2023</b>			
<ul style="list-style-type: none"> <li>Unidad de aprendizaje 4: Técnicas de Machine Learning en Big Data</li> </ul>	<ul style="list-style-type: none"> <li>Machine Learning para Big Data.</li> </ul>		
<b>Semana 13: del 12/06/2023 al 17/06/2023</b>			
<ul style="list-style-type: none"> <li>Unidad de aprendizaje 4: Técnicas de Machine Learning en Big Data</li> </ul>	<ul style="list-style-type: none"> <li>ML en Big Data - Supervisado</li> <li>ML en Big Data - No Supervisado</li> <li>Aplicaciones.</li> </ul> <b>PC2</b>		<ul style="list-style-type: none"> <li>Nota Trabajos</li> </ul>
<b>Semana 14: del 19/06/2023 al 24/06/2023</b>			

Unidades de aprendizaje	Contenidos y actividades a realizar	Recursos y materiales	Evaluaciones
<ul style="list-style-type: none"> <li>Unidad de aprendizaje 4: Técnicas de Machine Learning en Big Data</li> </ul>	<ul style="list-style-type: none"> <li>PySpark - ML</li> <li>Deep Learning para Big Data</li> <li>Aplicaciones.</li> </ul>		
<b>Semana 15 con feriado jueves 29: del 26/06/2023 al 01/07/2023</b>			
<ul style="list-style-type: none"> <li>Unidad de aprendizaje 4: Técnicas de Machine Learning en Big Data</li> </ul>	<ul style="list-style-type: none"> <li>Casos DL en Big Data</li> <li>Programación.</li> </ul> <p><b>Presentación Final de Proyectos</b></p>		<ul style="list-style-type: none"> <li>Nota Trabajos</li> </ul>
<b>Semana 16 de exámenes finales: del 03/07/2023 al 08/07/2023</b>			
<ul style="list-style-type: none"> <li>Unidad de aprendizaje 3: Plataformas de Big Data II</li> <li>Unidad de aprendizaje 4: Técnicas de Machine Learning en Big Data</li> </ul>	<b>Examen Final</b>		<ul style="list-style-type: none"> <li>Examen Final</li> </ul>

## VIII. Indicaciones para el desarrollo del curso

### **Plagio**

Se define como plagio según la RAE Copiar en lo sustancial obras ajenas, dándolas como propias. Esta definición no se limita a sólo los exámenes sino trabajos o todo ejercicio, resumen, mapa conceptual u otro que se solicite a los alumnos. En relación a los reglamentos vigentes, la sanción en estos casos es de obtener la calificación de cero y el reporte a las instancias correspondientes de la universidad para su respectiva suspensión o retiro de la universidad.

### **Entrega de trabajos**

Sólo se calificarán los trabajos entregados mediante Blackboard, los enviados por correo electrónico, físico u otro medio no serán considerados así hayan sido enviados dentro del plazo establecido para la entrega del trabajo.

### **Asistencia**

Como indica el actual el Reglamento de Estudios de Pregrado, la asistencia de los alumnos no es obligatoria en este curso y no será considerada como parte de la nota final. Sin embargo, tenemos Trabajos Prácticos Calificados y las fechas del cronograma son referenciales y pueden variar según el avance del curso. En ese sentido, se sugiere a los estudiantes estar presentes en todas las clases.

## IX. Referencias bibliográficas

### **Obligatoria**

Chambers, B. & Zahaira, M. (2018). *Spark: the definitive guide: big data processing made simple*. Sebastopol: O'Reilly Media, Inc.

White, T. (2015). *Hadoop: The Definitive Guide, Fourth Edition*. Sebastopol, CA: O'Reilly.