

MODALIDAD ACADÉMICA

Asignatura	Big Data: Arquitecturas y Estrategias para el Análisis de Datos Masivos	
Carrera	INGENIERÍA EN SISTEMAS DE INFORMACIÓN	
Ciclo Lectivo	2017	
Vigencia del programa	Desde el ciclo lectivo 2015	
Plan	2008	
Nivel	<input type="checkbox"/> 1er. Nivel <input type="checkbox"/> 2do. Nivel <input type="checkbox"/> 3er. Nivel <input type="checkbox"/> 4to. Nivel <input checked="" type="checkbox"/> 5to. Nivel	
Coordinador de la Cátedra	DEA Ing. Calixto Alejandro Maldonado	
Área de Conocimiento	<input checked="" type="checkbox"/> Programación <input checked="" type="checkbox"/> Computación <input checked="" type="checkbox"/> Sistemas de Información <input checked="" type="checkbox"/> Gestión Ingenieril <input type="checkbox"/> Modelos <input type="checkbox"/> Complementaria	
Carga horaria semanal	6 horas	
Anual/ cuatrimestral	cuatrimestral	
Contenidos Mínimos, según Diseño Curricular-Ordenanza 1150 (sólo para asignaturas curriculares)	Es una materia Electiva	
Correlativas para cursar (según Diseño Curricular-Ordenanza 1150)	Regulares	Aprobadas
	<ul style="list-style-type: none"> Al menos una materia del conjunto de PAV2 y DLC 	<ul style="list-style-type: none"> Diseño de Sistemas
Correlativas para rendir (según Diseño Curricular-Ordenanza 1150)	Regulares	Aprobadas
	<ul style="list-style-type: none"> Al menos una materia del conjunto de PAV2 y DLC 	<ul style="list-style-type: none"> Diseño de Sistemas
Objetivos de la Asignatura	<p>Esta materia le brindará al asistente los elementos teóricos y prácticos para poder agregar una nueva competencia a la altura de las demandas de la Industria de la tecnología de tratamiento de datos masivos porque lo hará capaz de extraer conocimiento útil de los datos en un contexto de proliferación de la producción de datos en las organizaciones y en la Red en general.</p> <p>Se espera que el estudiante, al finalizar la materia, pueda:</p> <ul style="list-style-type: none"> Conocer los conceptos básicos y la evolución histórica de la ciencia de datos, Big Data. Aplicar los conceptos asociados a la generación masiva de información proveniente de distintos orígenes. Entender las estructuras físicas y lógicas de una o más soluciones de Datos Masivos para poder elegir una a aplicar para resolver un caso práctico. Usar las soluciones de software disponibles en el mercado en aplicaciones a 	

	<p>problemas concretos de las mismas.</p> <ul style="list-style-type: none"> • Implementar una solución en el laboratorio para resolver un caso práctico. • Aplicar el proceso de Captación de datos masivos • Analizar, diseñar y programar una solución de datos masivos en un caso de la vida real. <p>Comprender las técnicas para la administración masiva de datos.</p>
<p><u>Programa Analítico</u></p>	
<p><u>Unidad Nro. 1: Características, Usos y Arquitecturas de BigData</u></p>	
<p>Objetivos Específicos: El asistente a la materia podrá en esta unidad:</p> <ul style="list-style-type: none"> • Conocer los conceptos principales y la evolución histórica de la Ciencia de Datos, Big Data. • Conocer los usos actuales y potenciales de BigData • Estrategias del proyecto de Big Data <p>Contenidos:</p> <ul style="list-style-type: none"> • Introducción a BigData. Conceptos y evolución histórica de la Ciencia de Datos, Big Data. • Usos de BigData. Retos y oportunidades. • Tecnologías Asociadas a BigData. • Open Data 	
<p><u>Bibliografía:</u></p> <p><i>Big data -Análisis de grandes volúmenes de datos en organizaciones- Joyanes Aguilar-Ed Alfaomega-ISBN: 978-607-707-689-6</i></p> <p><i>Big Data – La revolución de datos masivos – 2013 ISBN:978-84-15832-10-2</i></p>	
<p>Evaluación: Parcial Teórico 1 en la semana 8</p>	
<p><u>Unidad Nro. 2: Herramientas de BigData. Hadoop y otras.</u></p>	
<p>Objetivos Específicos: El asistente a la materia podrá en esta unidad:</p> <ul style="list-style-type: none"> • Transmitir sobre la diversidad de herramientas disponibles en el mercado • Utilizar la herramienta de BigData Hadoop aplicada en casos previamente enunciados. • Conocer la herramienta Vértica y ejemplos de su aplicación <p>Contenidos:</p> <ul style="list-style-type: none"> • Herramientas BigData. Hadoop <ul style="list-style-type: none"> • Técnicas de la administración masiva de datos • Obtención y limpieza de datos masivos • Programación bajo el Framework Hadoop • Minería de datos masivos (Inferencia estadística, modelos de regresión, clustering, árboles de decisión, K-nearest, Reglas de asociación, predicciones, otros) • Herramientas Big Data: Vertica, base de datos columnar con soporte completo de SQL <ul style="list-style-type: none"> • Arquitectura • Componentes • Ejemplos de uso 	
<p>Actividades: Actividad práctica integral. Diseño, desarrollo y visualización de una aplicación Big Data</p>	
<p><u>Bibliografía:</u></p>	

“Hernández Orallo, Ramirez Quintana, Ferri Ramirez. “Introducción a la Minería de Datos”
Editorial Pearson – Prentice Hall. 2004

Tom white “Hadoop The definitive Guide” 3° edición – oReilly - ISBN: 978-1-449-31152-0

Documentación oficial de HP-Vertica. Introducción y Herramientas

Big data -Análisis de grandes volúmenes de datos en organizaciones- Joyanes Aguilar-Ed Alfaomega-ISBN: 978-607-707-689-6

Evaluación: En la plataforma Virtual, Journals, Cuestionarios y Participacion en el foro.

Unidad Nro. 3: Bases de datos NoSQL

Objetivos Específicos: El asistente a la materia podrá en esta unidad:

- Comprender las diferentes tecnologías de almacenamiento escalables NoSQL y saber seleccionarlas y aplicarlas.

Contenidos:

- Bases de datos NonSQL. MongoDB, Cassandra y Apache Impala.
- Arquitectura de bases de datos NonSQL.
- Propiedades y Características.
- Usos de bases de datos NonSQL.
- Tecnologías asociadas a Bases de datos NonSQL.
- Vértica, base de datos columnar con soporte completo de SQL. Arquitectura, componentes y Ejemplos de uso.

Bibliografía:

Pramodkumar J. Sadalage, Martin Fowler “NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence”

Kristina Chodorow; MongoDB: The Definitive Guide; ISBN-13: 978-1449344689

Eben Hewitt; Cassandra: The Definitive Guide; ISBN-13: 978-1449390419

Evaluación: Trabajo Práctico Integral en la semana 15

Unidad Nro. 4: VISUALIZACIÓN DE CONOCIMIENTO

Objetivos Específicos: El asistente a la materia podrá en esta unidad:

- Comprender las técnicas para la visualización del conocimiento
- Adquirir competencias en desarrollo de aplicaciones de visualización de datos y aplicar el conocimiento en el trabajo práctico integral

Contenidos:

- Descubrimiento del conocimiento
- Desarrollo de productos de conocimiento
- Analítica de datos
- Visualización de datos

- Herramientas para la visualización de datos
- DataStoryTelling

Bibliografía:

- Robert D. Schneider “Hadoop For IBM Platform Computing” –ISBN: 978-1-118-25051-8
- Understanding Big Data – Analytics for Enterprise Class Hadoop and Streaming Data – IBM ISBN: 978-1-118-25051-8
- Tom white “Hadoop The definitive Guide” 3° edición – oReilly – ISBN: 978-1-449-31152-0
- Jason Venner “Pro Hadoop” - ISBN: 978-1-4302-1942-2
- Big data -Análisis de grandes volúmenes de datos en organizaciones- Joyanes Aguilar-Ed Alfaomega- ISBN: 978-607-707-689-6

Evaluación: Trabajo Práctico Integral en la semana 15

Metodología de enseñanza y aprendizaje	Exposición dialogada de los temas teóricos. Exhibición de Videos sobre la temática y debate posterior. Investigación de bibliografía recomendada. Práctica de Diseño de Datos, Instalación, Configuración, Carga y aplicación de Estrategias de análisis de datos masivos. Contacto con expertos del Medio para intercambiar experiencias en el campo. Presentación de trabajos prácticos de estudiantes realizados en años anteriores.																																	
Sistema de evaluación (Nombrar y describir cada una de las diferentes instancias de evaluación)	Se evaluarán la participación en los cuestionarios, diarios o journals y en Foros creados y publicados en la plataforma virtual de la materia en Moodle. Todos ellos tienen una evaluación y se aprueban con un puntaje igual o mayor que 4 (cuatro). Además deben aprobar la realización de un proyecto integrador de un caso con aplicación de las infraestructuras de datos vistas en la materia, previa aprobación del docente sobre la estrategia elegida por el grupo.																																	
Regularidad: condiciones (Describir las condiciones necesarias para regularizar. Se sugiere incluir la aclaración que el estudiante en condición de regular puede rendir en el plazo de un ciclo lectivo sin control de correlativas aprobadas)	<p>Las actividades evaluativas (AE) en la plataforma con 4 (cuatro). Escala de notas de regularidad(*)</p> <table border="1" data-bbox="678 1087 1247 1465"> <thead> <tr> <th>NOTAS</th> <th>PORCENTAJE</th> <th>CALIFICACIÓN</th> </tr> </thead> <tbody> <tr> <td>1</td> <td></td> <td>No Aprobado</td> </tr> <tr> <td>2</td> <td></td> <td>No Aprobado</td> </tr> <tr> <td>3</td> <td></td> <td>No Aprobado</td> </tr> <tr> <td>4</td> <td>55% a 57%</td> <td>Aprobado</td> </tr> <tr> <td>5</td> <td>58% a 59%</td> <td>Aprobado</td> </tr> <tr> <td>6</td> <td>60% a 68%</td> <td>Aprobado</td> </tr> <tr> <td>7</td> <td>69% a 77%</td> <td>Aprobado</td> </tr> <tr> <td>8</td> <td>78% a 86%</td> <td>Aprobado</td> </tr> <tr> <td>9</td> <td>87% a 95%</td> <td>Aprobado</td> </tr> <tr> <td>10</td> <td>96% a 100%</td> <td>Aprobado</td> </tr> </tbody> </table> <p>(*) Escala acordada en reunión de Docentes Coordinadores</p>	NOTAS	PORCENTAJE	CALIFICACIÓN	1		No Aprobado	2		No Aprobado	3		No Aprobado	4	55% a 57%	Aprobado	5	58% a 59%	Aprobado	6	60% a 68%	Aprobado	7	69% a 77%	Aprobado	8	78% a 86%	Aprobado	9	87% a 95%	Aprobado	10	96% a 100%	Aprobado
NOTAS	PORCENTAJE	CALIFICACIÓN																																
1		No Aprobado																																
2		No Aprobado																																
3		No Aprobado																																
4	55% a 57%	Aprobado																																
5	58% a 59%	Aprobado																																
6	60% a 68%	Aprobado																																
7	69% a 77%	Aprobado																																
8	78% a 86%	Aprobado																																
9	87% a 95%	Aprobado																																
10	96% a 100%	Aprobado																																
Promoción: condiciones (Aclarar si hubiera promoción de alguna parte de la asignatura, las condiciones y si tiene duración, con el mayor detalle posible)	No corresponde																																	
Aprobación Directa: condiciones. (la calificación será la nota registrada como Nota Final en Autogestión)	Se prevé otorgar la Aprobación Directa, si se aprueban los Actividades publicadas en la plataforma y el Proyecto integrador con nota 7 o más.																																	

(Se sugiere incluir la aclaración que el estudiante, en esta condición, puede registrar su nota en examen en el plazo de un ciclo lectivo, sin control de correlativas aprobadas, y después de ello se le exigirán correlativas aprobadas)																																		
<p>Modalidad de examen final (Describir las características metodológicas del examen final para los distintos estados del estudiante)</p>	<table border="1" data-bbox="678 527 1243 905"> <thead> <tr> <th>NOTA</th> <th>PORCENTAJE</th> <th>CALIFICACIÓN</th> </tr> </thead> <tbody> <tr><td>1</td><td></td><td>Insuficiente</td></tr> <tr><td>2</td><td></td><td>Insuficiente</td></tr> <tr><td>3</td><td></td><td>Insuficiente</td></tr> <tr><td>4</td><td></td><td>Insuficiente</td></tr> <tr><td>5</td><td></td><td>Insuficiente</td></tr> <tr><td>6</td><td>60% a 68%</td><td>Aprobado</td></tr> <tr><td>7</td><td>69% a 77%</td><td>Bueno</td></tr> <tr><td>8</td><td>78% a 86%</td><td>Muy Bueno</td></tr> <tr><td>9</td><td>87% a 95%</td><td>Distinguido</td></tr> <tr><td>10</td><td>96% a 100%</td><td>Sobresaliente</td></tr> </tbody> </table> <p data-bbox="586 915 1497 1037">Escala de Notas para Examen Final (*) Examen teórico oral, con sorteo de temas para exponer y confección de grilla de temas tomados y resultados parciales de cada tema para componer así la nota final.</p> <p data-bbox="586 1482 1240 1514">(*) Escala acordada en reunión de Docentes Coordinadores</p>	NOTA	PORCENTAJE	CALIFICACIÓN	1		Insuficiente	2		Insuficiente	3		Insuficiente	4		Insuficiente	5		Insuficiente	6	60% a 68%	Aprobado	7	69% a 77%	Bueno	8	78% a 86%	Muy Bueno	9	87% a 95%	Distinguido	10	96% a 100%	Sobresaliente
NOTA	PORCENTAJE	CALIFICACIÓN																																
1		Insuficiente																																
2		Insuficiente																																
3		Insuficiente																																
4		Insuficiente																																
5		Insuficiente																																
6	60% a 68%	Aprobado																																
7	69% a 77%	Bueno																																
8	78% a 86%	Muy Bueno																																
9	87% a 95%	Distinguido																																
10	96% a 100%	Sobresaliente																																
Actividades en laboratorio	Uso del Laboratorio IBM para Instalación, Configuración, Carga y Aplicación de Estrategias de análisis de datos masivos.																																	
Horas/año totales de la asignatura (hs. cátedra)	6 horas cátedra semanales por 15 semanas, resultando 90 horas totales																																	
Cantidad de horas prácticas totales (hs. cátedra)	45 horas de clases prácticas																																	
Cantidad de horas teóricas totales (hs. cátedra)	6 horas cátedra semanales																																	
Tipo de formación práctica (sólo si es asignatura curricular -no electiva-)	<input type="checkbox"/> Formación experimental <input type="checkbox"/> Resolución de problemas de ingeniería <input type="checkbox"/> Actividades de proyecto y diseño <input type="checkbox"/> Prácticas supervisadas en los sectores productivos y /o de servicios																																	

<p>Cantidad de horas cátedras afectadas a la formación práctica indicada en el punto anterior (sólo si es asignatura curricular -no electiva-)</p>	<p>No aplica</p>
<p>Descripción de los prácticos</p>	<p>Se realizará un trabajo práctico integrador grupal. Desarrollo de un proyecto Big Data que permitirá a los estudiantes recorrer el ciclo de vida completo desde la planificación, análisis, la programación de la aplicación y el desarrollo de una interfaz de visualización del conocimiento Big Data.</p>
<p>Criterios generales (los cuales serán tenidos en cuenta en las correcciones)</p>	<p>Evaluación del proceso:</p> <ul style="list-style-type: none"> - Colaboración grupal e inter grupal - Interés - Liderazgo y motivación - Comunicación grupal y con el docente - Construcción del conocimiento - Creatividad - Aplicación del conocimiento de otras asignaturas <p>Evaluación del trabajo presentado:</p> <ul style="list-style-type: none"> - Aspectos formales - Puntualidad de entrega - Claridad - Innovación - Funcionamiento de la aplicación desarrollada <p>El trabajo práctico una vez finalizado deberá ser expuesto y defendido en una presentación oral.</p> <p>Criterios de la presentación oral:</p> <ul style="list-style-type: none"> • Expresión Oral. • Dominio del lenguaje técnico apropiado. • Conocimiento del dominio objeto de la Actividad Práctica. • Conocimiento de las actividades realizadas. • Conocimiento del método y las herramientas para el desarrollo de la Actividad Práctica. • Experiencias aportadas por la Administración del Proyecto. <p>Conclusiones personales de la realización de la Actividad Práctica</p>

Cronograma de actividades de la asignatura (contemplando las fechas del calendario 2017 y para cada unidad)	Sem N°	Unid N°	HsTeóricas	HsPráct	Contenido
	1	1 y 2	3	3	Características, Usos y Arquitecturas de BigData. Herramientas de BigData
	2		3	3	Arquitectura y Estrategias
	3		3	3	Presentación Hadoop. Trabajo Práctico, presentación
	4		3	3	Hadoop - Carga de Datos y primeros pasos. Minería de Datos Masivos -
	5		3	3	Hadoop Ejemplos de uso
	6		3	3	Vertica Arquitectura y herramientas
	7		3	3	Vertica - Reunión Control de TP
	8		3	3	Atención de consultas - Parcial Teórico 1
	9	3	3	3	Bases de datos NoSQL Propiedades y Características
	10		3	3	MongoDB - Reunión Control de TP
	11		3	3	Cassandra y Apache Impala
	12		3	3	Cassandra - Apache Impala - Reunión Control de TP
	13	4	3	3	Visualización de Conocimiento y Datos
	14		3	3	Herramientas de Visualización - Presentación de TP
15		3	3	Revisión de aceptación de Trabajo Práctico	
Propuesta para la atención de consultas y mail de contacto.	Se prevé una sesión de inicio del Trabajo Practico y tres instancias de control de avance (semanas 7, 10 y 12) y antes del Parcial Teórico, una instancia de consultas sobre los temas teóricos (semana 8). Se construirá el material de soporte de clases para poner a disposición en el Campus Virtual http://uv.frc.utn.edu.ar . Se proporcionarán las direcciones de mail de los profesores para consultas ad-hoc				
Plan de integración con otras asignaturas	La materia tiene base en Gestión de Datos y Sistemas Operativos, se realizarán actividades de revisión de los contenidos relacionados, como Arquitectura de Motores de Bases de Datos y Modelado de Datos y Normalización de Gestión de Datos. Con sistemas Operativos, File Systems y Administración de Memoria.				
Bibliografía Obligatoria	<ul style="list-style-type: none"> Joyanes Aguilar – Big Data: Análisis de Grandes volúmenes de datos en organizaciones. AlfaOmega ISBN: 978-607-707-689-6 Viktor Mayer-Schönberger, Kenneth Cukier “Big Data – La revolución de datos masivos” – 2013 ISBN: 978-84-15832-10-2 Tom White “Hadoop The definitive Guide” 3° ed. – O’Reilly – ISBN: 978-1-449-31152-0 Martin Fowler, Pramodkumar J. Sadalage “NoSQL Distilled: A Brief Guide to the Emerging World of PolyglotPersistence” Kristina Chodorow “MongoDB: The Definitive Guide”; ISBN-13: 978-1449344689 Eben Hewitt; “Cassandra: The Definitive Guide”; ISBN-13: 978- 				

	<p>1449390419</p> <ul style="list-style-type: none"> • Hernández Orallo, Ramirez Quintana, Ferri Ramirez. “Introducción a la Minería de Datos” Editorial Pearson – Prentice Hall. 2004
<p>Bibliografía Complementaria</p>	<ul style="list-style-type: none"> • Mike Barlow “Real-Time Big Data Analytics: Emerging Architecture” • Raul F. Chong and Clara Liu, “DB2 Essentials: Understanding DB2 in a Big Data World” (3rd Edition) (IBM Press) (Nov 9, 2013) • Thomas H. Davenport, Big Data at Work: Dispelling the Myths, Uncovering the Opportunities • Mike Barlow The Culture of Big Data • Hurwitz , Alan Nugent , Fern Halper , Marcia Kaufman “Big Data For Dummies” Paperback • Lars Nielsen ”The Little Book of Cloud Computing, 2013 Edition: Including Coverage of Big Data Tools” • David Feinleib ”Big Data Demystified: How Big Data Is Changing The Way We Live, Love And Learn” • O’Reilly Big Data Now: 2012 Edition O’Reilly Media Inc. • Thomas H. Davenport “At the Big Data Crossroads” – Amdeus- 2013 • Robert D. Schneider, “Hadoop For IBM Platform Computing”–ISBN: 978-1-118-25051-8 • Jason Venner Pro Hadoop - - ISBN: 978-1-4302-1942-2 • Zikopolous Paul, Deroos Dirk, Deutsch Tom, Lapis George, Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, McGraw-Hill, 2012 • Foster Kevin, Nathan Senthil, Rajan Deepak, Ballard Chuck, IBM InfoSphere Streams: Assembling Continuous Insight in the Information Revolution, IBM RedBooks, 2011 • Cisco, “Internet será cuatro veces más grande en 2016”, Artículo Web http://www.cisco.com/web/ES/about/press/2012/2012-05-30-internet-sera-cuatro-veces-mas-grande-en-2016--informe-vini-de-cisco.html • Soares Sunil, “Not Your Type? Big Data Matchmaker On Five Data Types You Need To Explore Today”, Artículo Web http://www.dataversity.net/not-your-type-big-data-matchmaker-on-five-data-types-you-need-to-explore-today/ • Clegg Dai, Big Data: “The Data Velocity Discussion”, Artículo Web http://thinking.netezza.com/blog/big-data-data-velocity-discussion • Kobielus James, “Big Data Analytics Helps Researchers Drill Deeper into Multiple Sclerosis”, Artículo Web http://thinking.netezza.com/blog/big-data-analytics-helps-researchers-drill-deeper-multiple-sclerosis • Biesdorf, Court, Willmott “Big data:What’s your plan?” McKinsey Quartely- – 2013 • McKinsey “Making data analytics work” – 2013

Distribución de docentes	<i>Curso</i>	<i>Turno</i>	<i>Día y Horas</i>	<i>Profesor</i>	<i>JTP</i>	<i>Ayudante</i>
	5k_5	Noche	Miercoles de 18 a 20	DEA Ing. Calixto Maldonado	Ing. Pablo Vaca	
	5k_5	Noche	Viernes de 20 a 22:30	Mag. Franco Mana	Ing. Pablo Vaca	

Firma:

Aclaración: