

Relatores:

Wenceslao Palma. Dr en Informática, Universidad de Nantes, Francia. Docente e investigador de la Escuela de Ingeniería Informática PUCV en el área de Big Data.

Héctor Allende-Cid. Dr en Ingeniería Informática, Universidad Técnica Federico Santa María, Chile. Docente e investigador de la Escuela de Ingeniería Informática PUCV en el área de Data Science.

Rodrigo Alfaro. Dr (c) en Ingeniería Informática, Universidad Técnica Federico Santa María, Chile. Docente e investigador de la Escuela de Ingeniería Informática PUCV en el área de Data Science. Co-founder de Analytic, empresa dedicada a la recolección, clasificación y análisis de Big Data.

Charlas invitadas:

Se contará con expositores invitados para temas específicos.

Metodología:

Clases expositivas, actividades de laboratorio.

Duración:

81 horas.

Horario de clases:

Las clases se dictan en dos sesiones semanales de 3 hrs c/u, en los siguientes horarios: Viernes de 19:00 hrs. a 22:15 hrs y Sábado de 10:00 hrs. a 13:15 hrs.

Planificación de clases:

3,4,17,18,24 y 25 de Abril.
8,9,15,16,29 y 30 de Mayo.
5,6,12,13,19,20,26 y 27 de Junio.
3,4,10,11,24,25 y 31 de Julio.

Evaluación:

Al final de cada módulo cada relator realizará una evaluación final que será calificada en el rango 1,0 a 7,0.

Arancel:

\$1.700.000 general (\$1.400.000 si la inscripción se realiza antes del 24 de enero de 2020).
\$1.400.000 ex-alumnos, funcionarios PUCV.

Formas de pago:

Documentación en 5 cheques, tarjetas débito/crédito, transferencia electrónica.

Asistencia:

Se exige un mínimo de asistencia de 75% respecto de las 81 horas programadas.

Lugar de realización:

Centro de Estudios Avanzados y Extensión PUCV, Antonio Bellet #314, Providencia, Santiago.

Inscripciones y contacto

Email: bigdata@pucv.cl
Director: Dr. Wenceslao Palma
[wenceslao.palma@pucv.cl]



Santiago 2020 Diplomado en Big Data y Data Science

ESCUELA DE
INGENIERÍA INFORMÁTICA



PONTIFICIA
UNIVERSIDAD
CATÓLICA DE
VALPARAÍSO

Diplomado en Big Data y Data Science

Dirigido a profesionales del área informática o afín que requieran utilizar herramientas computacionales y estadísticas para procesar y analizar grandes volúmenes de datos.

Objetivos

- Comprender el contexto y las tendencias que permiten generar conocimiento explícito sobre la base del almacenamiento de grandes volúmenes de datos.
- Comprender los problemas asociados al procesamiento y análisis de grandes volúmenes de datos.
- Conocer y experimentar con modelos, algoritmos y herramientas asociados al procesamiento y análisis de grandes volúmenes de datos.
- Desarrollar aplicaciones que procesan y analizan grandes volúmenes de datos.

Contenidos

Módulo 1: Datos como fuente de valor

Introducción y contexto

Sociedad de la Información y calidad de vida.

La nueva era de la máquina.

Calidad de los datos y buenas prácticas.

Big Data en la organización

Big Data en la Gestión de Operaciones.

Big Data en la Gestión del Control.

Big Data en la Gestión Estratégica.

Problemas de negocio y soluciones basadas en datos

De problemas de negocios a tareas de minería de datos.

El proceso de minería de datos.

Nuevos modelos de negocio basados en Big Data.

Futuro del Big Data

Módulo 2: Big Data

Introducción

Procesamiento de grandes volúmenes de datos.

Mapreduce, NoSQL, flujos de datos.

Proveedores de servicios: cloud, hosting, ecosistemas.

Hadoop

Distribución de datos (HDFS), tolerancia a fallas y balance de carga.

El ecosistema Hadoop.

MapReduce.

Hbase, Hive.

Pig.

Spark.

Sistemas de Gestión de Flujos de Datos

Arquitectura.

Procesamiento de consultas. Operadores de Consulta. Lenguajes de Consulta.

Procesamiento distribuido de flujos de datos.

Spark.

Módulo 3: Data Science

Análisis de Datos

Estadística Descriptiva e Inferencia.

Análisis exploratorio de datos.

Identificación de relaciones entre variables.

Técnicas de Análisis Cuantitativo.

Visualización de datos.

Machine Learning para Big Data

Clasificación, Agrupamiento

y Reducción de dimensionalidad.

Apache Spark.

Machine Learning Library en Spark.

Enfoques Paralelos y Distribuidos en Machine Learning.

Reconocimiento de Patrones y Machine Learning

Aprendiendo de los datos.

Clasificación, Predicción y Agrupamiento.

Modelos Lineales.

Arboles de Clasificación.

Regresión Logística.

Redes Neuronales Artificiales.

Máquinas de Soporte Vectorial.

Algoritmo de K-medias.