



**Universidad  
Ricardo Palma  
Rectorado**

**Programa de Especialización en ciencia de Datos**

## **II TALLER DE ESPECIALIZACIÓN EN: "BIG DATA APLICADO"**

**(Aprobado con Acuerdo de Consejo Universitario No. 0911-2018)**

**(13, 15, 17, 20, 22, 24, 27, 29, 31 Agosto y  
03 de Setiembre del 2018)**



# **BIG DATA APLICADO**

### **Informes e inscripción:**

70 80 000 anexo 0169 – Piso 8 de la Torre Administrativa – Universidad Ricardo Palma, Av. Benavides 5440 – Santiago de Surco.

[destadistico@gmail.com](mailto:destadistico@gmail.com) / [oroque@urp.edu.pe](mailto:oroque@urp.edu.pe) / [bigdataplicado@gmail.com](mailto:bigdataplicado@gmail.com)

962 739 102

968 248 582



## PRESENTACIÓN

El Taller de Especialización en "BIG DATA" en su segunda versión trata de difundir de manera práctica la arquitectura, herramientas y tecnologías Big Data y su aplicación en la empresa. Se abarcará los fundamentos de Hadoop, taller de introducción a Python, fundamentos de Spark, Machine Learning y se realizará la implementación de un caso de estudio para que todos los participantes aprendan de manera práctica estas nuevas tecnologías.

El taller será monitoreado por expertos en el tema y con experiencias en este tipo de enseñanza y en esta ocasión se realizará en la Universidad Ricardo Palma. Se entregará el material y los códigos de las aplicaciones en Python y Pyspark, el cual ha sido elaborado por los expositores y constituye un material protegido con derecho de autor contra cualquier reproducción no autorizada.

La experiencia obtenida en el primer taller de especialización de Big Data, nos ha motivado plantear por segunda vez el taller de Big Data Aplicado, agregando 6 horas adicionales en el desarrollo del caso de estudio, el cual será 100% práctico para que los participantes puedan adquirir de mejor manera los conocimientos en Big Data.

Este taller de especialización se organiza dentro del marco de la Maestría en Ciencia de los Datos que se dicta en la Universidad Ricardo Palma.

## PROGRAMA TEMÁTICO

### PROGRAMA CURRICULAR INTEGRAL:

#### 1. INTRODUCCION AL BIG DATA

- Que es Big Data
- Pipeline Arquitectura Tradicional
- Pipeline Arquitectura Big Data
- Distribuciones Hadoop
- Infraestructura OnPremise – Cloud

#### 2. FUNDAMENTOS HADOOP

- Que es Apache Hadoop?
- Componentes de Hadoop
  - HDFS
  - MapReduce
  - Yarn
  - Common Utilities
- Ecosistema Hadoop

#### 3. TALLER DE INTRODUCCIÓN A PYTHON

- Introducción al Python
- Manejo de cadenas
- Manejo de Números
- Listas, tuplas y diccionarios
- Funciones
- Manejo de librerías Pandas.

#### 4. FUNDAMENTOS SPARK

- Introduction to Apache Spark
- Spark Driver, Workers process and cluster manager.
- RDDs, Daframes y Datasets

- RDDS operations
- SparkSQL
- Spark-submit for execute on the cluster

## 5. MACHINE LEARNING

- Introducción al Machine Learning
- Comandos principales para el manejo de archivos (entrenamiento, prueba, balanceo)
- Generación de modelos Random forest, Decision tree, Gradient boosted tree, Logistic regression, Multilayer perceptron (neural net), Naive Bayes
- Indicadores principales del modelo.

## 6. IMPLEMENTACIÓN DE UN CASO DE ESTUDIO USANDO CLOUDERA CDH

- **Serialización:** Avro y Parquet
- **Ingesta:** HDFS, Sqoop, PySpark
- **Procesamiento:** Hive, Impala, Pyspark
- **Analítica:** Pyspark + Machine Learning (MLib)
- **Monitoreo y Workflow:** Cloudera Manager, Hue, Oozie

## OBJETIVOS:

El participante adquirirá y aplicará:

- Conocimientos sobre la arquitectura, distribuciones y ecosistema Hadoop.
- Conocimientos básicos del lenguaje de programación Python.
- Conocimientos del ambiente de cómputo distribuido en el lenguaje Python.
- Técnicas de ingesta de datos con HDFS, Sqoop y Pyspark.
- Procesamiento distribuido de grandes volúmenes de datos (Pyspark).
- Conocimientos básicos de Machine Learning aplicados a un caso de Negocio.
- Técnicas Analíticas con Pyspark, Hive e Impala.

## FUNDAMENTACIÓN ECONÓMICA:

- Profesionales en general: S/. 1,500.00 = 2 participantes
- Miembro URP: S/. 1,400.00 = 6 participantes
- Estudiantes de Pregrado y Posgrado: S/. 1,300.00 = 5 participantes

### Corporativo y costo promocional

- **Si se inscriben 2 participantes en forma conjunta:** S/. 1,400.00
- **Si se inscriben 3 participantes o más en forma conjunta:** S/. 1,300.00  
**= 3 participantes (una promoción)**

## PERFIL DE LOS PARTICIPANTES:

- Tener conocimientos básicos de matemáticas, ciencia de la computación, y de estadística, en lo referente a la comprensión general de la terminología y los conceptos de cada área.
- Conocimiento básico de bases de datos.
- Conceptos generales de lenguajes de programación (Deseable Python).
- Concepto de Big Data Básicos (Deseable).
- Conceptos básicos de Machine Learning.
- Linux Básico (Deseable).
- Profesionales de las carreras de ingeniería informática, ingeniería de sistemas, ingeniería industrial, estadística y de otras carreras profesionales que estén interesados en conocer las herramientas y tecnologías de Big Data y su aplicación en un caso de uso de Negocio.
- Desarrolladores de sistemas, data scientist, analistas y programadores interesados en aprender, aplicar y/o implementar proyectos y herramientas de Big Data en su empresa.

## EXPOSITORES:

### ❖ Mg. Eder Pineda Claros

Ingeniero de Computación y Sistemas con gran capacidad de desempeño y enfocado a lograr soluciones inmediatas a los negocios utilizando herramientas de inteligencia de negocio y analíticas, desempeñándose actualmente como Coordinador de Big Data Analytics en Entel Perú.

- Maestrando de la Maestría en Ciencia de los Datos de la Universidad Ricardo Palma
- Magister Business Intelligence de la Universidad de Barcelona
- Arquitecto de integración en EGS Group
- Asesor en Pentaho Perú
- Docente de cursos en estadística para no estadísticos en la Universidad Ricardo Palma

### ❖ Ing. David Narváez

Ingeniero de Sistemas, Especialista en Big Data, Business Intelligence y Data Warehousing. Arquitecto de Soluciones y responsable de la gestión de proyectos de Big Data y BI para distintos tipos de industrias como: financiera, seguros, mercado de valores, académica y telecomunicaciones.

Actualmente desempeñándose como Coordinador de Big Data y Analytics en Entel Peru, responsable de la implementación de la plataforma tecnológica de Big Data y Casos de Uso relacionados con el negocio.

## METODOLOGÍA:

- Teórica y práctica en un laboratorio de cómputo.
- La enseñanza será desarrollada haciendo uso de Cloudera CDH, implementada en plataforma AWS (Amazon Web Services). Se hará uso del software Jupyter, el cual nos permitirá realizar la programación en Python.
- Uso de herramientas del ecosistema de Hadoop como Hive, Impala, Spark y Hue

- Asistencia presencial 100%.
- Asesoría y atención de consultas durante la duración del curso taller de especialización.
- Materiales diseñados por los expositores, el cual se les entregará a cada participante.
- Se dictará en un laboratorio de cómputo, una PC por participante.
- Horas adicionales de práctica de la plataforma Big Data, de manera que el alumno pueda reforzar sus conocimientos fuera de clase.
- 03 Evaluaciones teórico/práctico de los conocimientos adquiridos.

### CERTIFICACIÓN:

- Certificado por haber aprobado el **II Taller de Especialización en: "Big Data Aplicado"** expedido por el Rectorado de la Universidad Ricardo Palma.  
Para obtener el certificado deberá tener la nota mínima de catorce (14).

### REQUISITOS:

- Escaneado del comprobante de pago
- Ficha de inscripción
- Constancia de estudios superiores
- Copia del DNI
- 2 fotos tamaño carnet a color con fondo blanco

### LUGAR Y HORARIO:

**Lugar:** Laboratorio de Cómputo del Aulario Inca Roca la Universidad Ricardo Palma, ubicada en la Av. Benavides 5440 – Santiago de Surco.

**Horario:** Lunes-Miércoles-Viernes de 19:00 – 22:00 horas

