



### ● OBJETIVO GENERAL:

Brindar un panorama amplio sobre el emergente problema de bigdata así como aprender los fundamentos de una de las principales herramientas que solventa la problemática (Hadoop).

Por tal motivo se aborda la instalación en modo clúster, la programación de una primera aplicación que atiende un problema de bigdata y finalmente el monitoreo del clúster.

### ● DIRIGIDO A:

Profesionistas y estudiantes interesados en conocer los fundamentos de Hadoop y que quieran aprender a montar un ambiente de desarrollo.

### ● PERFIL DE INGRESO:

Conocimientos básicos en Linux, redes y programación Java.

## CONTENIDO

- BIGDATA Y LA NUBE
- ARQUITECTURAS EMERGENTES PARA BIGDATA
- HADOOP
- INSTALACIÓN DE HADOOP EN MODO CLÚSTER
- MAPREDUCE
- ECOSISTEMA DE HADOOP
- MONITOREO DEL CLÚSTER

## Aspectos relevantes:

- Los instructores asignados son personal calificado con experiencia y cuentan con estudios de Maestría y la mayoría han sido formados en los programas de posgrado que imparte este Centro de Investigación, con registro en el Padrón Nacional de Posgrados de Calidad (PNPC) del CONACYT.
- El Centro de Investigación en Computación del IPN realiza investigación científica, así como desarrollo de aplicaciones en las ciencias de la computación e ingeniería de cómputo.
- La Secretaría del Trabajo y Previsión Social–STPS le otorgó a este Centro la constancia como agente capacitador externo con número de registro IPN-811229-H26-0013 para impartir cursos de capacitación y adiestramiento señalados en la solicitud de fecha 15 de junio del 2011.
- La relación de cursos que están registrados puede consultarla en el archivo en PDF y puede descargarla en la siguiente liga. [www.cursos.cic.ipn.mx](http://www.cursos.cic.ipn.mx). El registro asignado a través de esta constancia no otorga valor curricular o académico a los programas y/o cursos autorizados.

## Secuencia académica

### CONTÁCTANOS

Dirección: Av. Juan de Dios Bátiz s/n  
casi esq. Miguel Othón de Mendizábal.  
Unidad Profesional Adolfo López Mateos

Col. Nueva Industrial Vallejo  
Delegación Gustavo A. Madero  
C.P 07738, Cd. de México.

De 8:30 a 14:00 hrs y de 17:00 a 20:00hrs  
Tel. 57.29.60.00, ext. 56529, 56605, 56622

**DDEP**  
Departamento de Diplomados  
y Extensión Profesional



### CURSO 1

**Hadoop, conceptos,  
instalación y  
primera aplicación**

## CONTENIDO TEMÁTICO

1. **BIGDATA Y LA NUBE**
2. **ARQUITECTURAS EMERGENTES PARA BIGDATA**
  - 2.1 Oracle, IBM, Pivotal, SAP, LexisNexis
  - 2.2 Demanda de alto procesamiento
3. **HADOOP**
  - 3.1 Arquitectura HDFS
  - 3.2 Arquitectura MapReduce
  - 3.3 Redireccionamiento de entrada/salida y descriptores de archivos
4. **INSTALACIÓN DE HADOOP EN MODO CLÚSTER**
  - 4.1 Requisitos de hardware
  - 4.2 Requisitos de software (Java, SSH, Linux, certificados digitales)
  - 4.3 core-site.xml, mapred-site.xml, hdfs-site.xml, yarn-site.xml
  - 4.4 Comprobación del nodo maestro
  - 4.5 Comprobación de los nodos esclavos
5. **MAPREDUCE**
  - 5.1 Problemáticas abordadas por MapReduce
  - 5.2 Programación de aplicación
  - 5.3 Ejecución y pruebas de la aplicación
6. **ECOSISTEMA DE HADOOP**
  - 6.1 Pig, Hive, HBase, Zookeeper
7. **MONITOREO DEL CLUSTER**