



● OBJETIVO GENERAL:

Comprender y utilizar herramientas que potencializan al clúster de Hadoop para hacerlo más robusto y para atender necesidades específicas de Big Data como pueden ser procesos automatizados de carga de datos, la consulta de datos semi-estructurados utilizando SQL o la programación de aplicaciones hechas a medida para la explotación de datos.

● DIRIGIDO A:

Profesionistas y estudiantes que enfrenten problemas emergentes relacionados con Big Data.

● PERFIL DE INGRESO:

Es requisito contar con: 1) conocimiento intermedio en Linux, 2) conocimientos intermedios en Java (SWING no es java intermedio), 3) conocimientos básicos de SQL y sobretodo 4) experiencia ya sea en la instalación o manipulación de un Clúster de Hadoop. Es muy difícil sobrellevar el curso sino se cumplen estos requisitos.stema

CONTENIDO

- HDFS y YARN
- CARGA MASIVA DE DATOS CON FLUME
- CONSULTA SQL CON HIVE
- PROCESAMIENTO DE DATOS CON SPARK

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. El registro asignado a través de esta constancia no otorga valor curricular o académico a los programas y/o cursos autorizados.

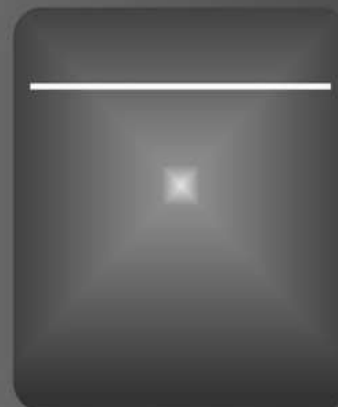
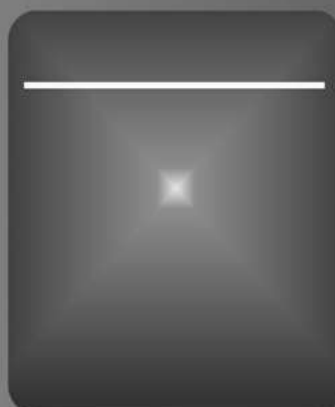
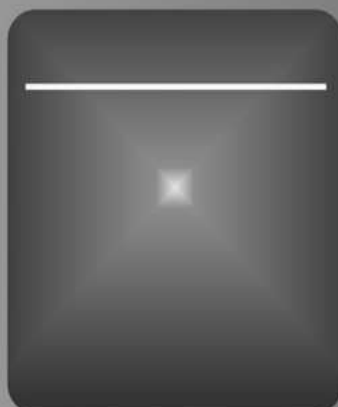
CONTÁCTANOS

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



CONTENIDO TEMÁTICO

1. HDFS y YARN
 - 1.1 Arquitectura HDFS
 - 1.2 Arquitectura YARN
 - 1.3 Instalación del Clúster

2. CARGA MASIVA DE DATOS CON FLUME
 - 2.1 Arquitectura e instalación
 - 2.2 Sources y Sinks
 - 2.3 Stream y Files

3. HADOOP
 - 3.1 Arquitectura HDFS
 - 3.2 Arquitectura MapReduce
 - 3.3 Redireccionamiento de entrada/salida y descriptores de archivos

4. CONSULTA SQL CON HIVE
 - 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. PROCESAMIENTO DE DATOS CON SPARK
 - 5.1 Arquitectura e instalación
 - 5.2 RDDs
 - 5.3 Spark Core