



Universidad Nacional de Luján
REPUBLICA ARGENTINA



Nº DISPOSICION:

Nombre de la actividad: Taller de Capacitación en Hadoop y tecnologías aplicadas para Big Data.

Descripción de la actividad: Taller de Capacitación a todo el personal de la Dirección General de Sistemas sobre el framework de desarrollo de aplicaciones Hadoop.

GRUPO DE TRABAJO: Personal Técnico de la Dirección General de Sistemas (DGS): • A.S. Pablo Tomas Delvechio	Cantidad Total de Horas: 10Hs. Horas Teoría: 4Hs. Horas Práctica: 6Hs.
FUNDAMENTACIÓN: El Big Data es un enfoque de análisis y procesamiento de datos que está adquiriendo cada vez más relevancia en los procesos de diferentes organizaciones. Las posibilidades de explotar la información disponible dependen de los intereses de las organizaciones y del uso adecuado de las herramientas existentes. En los últimos años, se impuso en la industria y la academia la herramienta Hadoop como estándar de facto. Esta herramienta es libre, y permite la construcción de un cluster de procesamiento masivo con muy bajos costos y altas prestaciones. Resulta de interés que los desarrolladores tengan un conocimiento de la herramienta para empezar a pensar posibles aplicaciones concretas en el ámbito de la Universidad.	
OBJETIVOS GENERALES: • Exponer conceptos básicos de desarrollo de aplicaciones distribuidas. • Ofrecer una introducción al enfoque MapReduce de desarrollo y su arquitectura. • Instalar un cluster Hadoop en modo desarrollo. • Desarrollar aplicaciones y pruebas de ejecución.	
MODALIDAD: • PRESENCIAL • Se realizara una exposición de los conceptos principales. • Se ofrecerá un espacio de intercambio de dudas. • Se procederá a realizar trabajos en los equipos informáticos disponibles. El trabajo práctico consistirá en la instalación de Hadoop y el desarrollo de una aplicación según las reglas indicadas por el expositor. • Se realizará una prueba de ejecución en cada instalación.	

DESTINATARIOS:

Todo el personal de la DGS.

CANTIDAD MÁXIMA DE ALUMNOS:

20 personas

REQUISITOS PARA EL CURSADO:

Conocimientos básicos de desarrollo de software. Manejo básico de Sistema Operativo GNU/Linux.

CANTIDAD DE ENCUENTROS:

La capacitación está organizada en 4 encuentros de 2 horas y media cada uno. El primer encuentro es introducción a los conceptos principales de la plataforma. Los encuentros siguientes se destinarán al desarrollo de las habilidades prácticas.

CERTIFICADOS QUE SE OTORGAN Y CONDICIONES:

Certificado de ASISTENCIA:

- Haber asistido al 80% de los encuentros.

Certificado de APROBACION:

- Haber asistido al 80% de los encuentros.
- Ejecutar una aplicación desarrollada en la instalación Hadoop.

RECURSOS NECESARIOS:

- Se deberá disponer de un espacio físico para el dictado de la Capacitación.
- Cada participante debe contar con un equipo informático (propio o de la institución) con Sistema GNU/Linux (Debian o Ubuntu) y Conexión a Internet.
- Proyector para diapositivas.

Unidad I

Objetivos:

Introducción al desarrollo de aplicaciones distribuidas.

Desarrollo:

- Desarrollo centralizado vs distribuido.
- Arquitecturas de sistemas distribuido.
- Teorema CAP

Unidad II

Objetivos:

Introducción a Hadoop.

Desarrollo:

- Introducción a Hadoop
- Diferencias entre Hadoop 1 y 2
- Arquitectura de HDFS
- Arquitectura de YARN

Unidad III

Objetivos:

Instalación de un cluster Hadoop

Desarrollo:

- Instalación de un cluster.
- Diferencias entre una instalación de desarrollo y una instalación de producción.
- Configuración de una instalación Hadoop.
- Optimizaciones y problemas comunes.

Unidad IV

Objetivos:

Desarrollo de aplicaciones en Hadoop

Desarrollo:

- Presentación de ejemplos de aplicaciones MapReduce
- Utilización de HDFS
- API de Hadoop
- Desarrollo de un programa en código JAVA.
- Ejecución de una aplicación MapReduce.

BIBLIOGRAFIA

- [DEA,2004]: Dean, J. Et. all. "*MapReduce: Simplified Data Processing on Large Clusters*". OSDI. 2004.
- [SHV,2010]: Shvachko, K. Et. all. "*The Hadoop Distributed File System*". IEEE. 2010.
- [WHI,2012]: T. White, *Hadoop: The definitive guide*, 2012.

100

Relecho

Relecho Pablo Tomas

Shvachko