

INFRAESTRUCTURA PARA BIG DATA

Nº de Créditos: **6 ECTS**
Primer Semestre
Segundo Curso

EQUIPO DOCENTE

Gómez Arribas, Francisco
Escuela Politécnica Superior
UAM
Coordinador

García Dorado, José Luis
Escuela Politécnica Superior
UAM

González Martínez, Iván
Escuela Politécnica Superior
UAM

Mora Rincón, Miguel Ángel
Escuela Politécnica Superior
UAM

Moreno Martínez, Víctor
Escuela Politécnica Superior
UAM

Ordax Cassa, José Miguel
Arquitecto de Tecnologías de
Información
IBM

Ramos de Santiago, Javier
Escuela Politécnica Superior
UAM

Riesco García, Diego
Arquitecto de Tecnologías de
Información
IBM

Villegas Núñez, Paulo
Escuela Politécnica Superior
UAM

OBJETIVOS

- Configurar sistemas multicore y multicpu para su utilización con MapReduce y Hadoop
- Definir los recursos necesarios para abordar aplicaciones que almacenan y procesan grandes volúmenes de datos, evaluando la idoneidad de su implementación en plataformas de computación y comunicación de altas de prestaciones.

- Analizar los requisitos para una implementación eficaz de las aplicaciones de Big Data y de las características deseables en las arquitecturas donde se ejecutan.
- Instalar y administrar un cluster Hadoop y puesta en marcha del ecosistema de aplicaciones asociado.
- Utilizar la plataforma Hadoop así como las diversas herramientas del ecosistema Hadoop.
- Instalar y desplegar un Cloud privado, administrar despliegues en el Cloud y monitorizar despliegues en el Cloud público
- Entender qué significan los modelos de Plataforma como Servicio y cómo pueden ayudar a agilizar los procesos de aprovisionamiento y consumo de servicios, entender cómo los modelos de PaaS permiten mejorar el "time to market" en los ciclos de desarrollo de aplicaciones tanto en pequeñas empresas y start-ups, como en grandes organizaciones.
- Utilizar de forma práctica una plataforma PaaS (Bluemix), trabajando de forma práctica en la creación de aplicaciones consumiendo los servicios que se ofrecen y su aplicación a potenciales casos reales.

PROGRAMA DETALLADO

- Arquitecturas para tratar grandes volúmenes de información
 - Arquitecturas de referencia para Hadoop
 - Instalación y configuración de un cluster Hadoop
 - Uso de herramientas de planificación y gestión.
 - Hadoop BigInsights
- Supervisión y mantenimiento de un cluster Hadoop
 - Estado de HDFS y copia de datos
 - Añadir y quitar nodos
 - Balanceo del cluster
 - Impacto de la red de comunicación en un cluster Big Data.
 - Copias de seguridad
- Evaluación de prestaciones y optimización en un caso práctico
 - Benchmarking y tuneado de parámetros
 - Caso práctico: procesamiento Big Data de datos de red
- Infraestructura para otros entornos Big Data: Ecosistema Spark
 - Arquitectura de un sistema Spark. Estructura interna y flujos de datos
 - Modos de ejecución en Spark: local vs cluster. Ámbitos y contextos: driver, ejecutores

- Gestión de memoria
- Modos de ejecución en cluster. Comunicación entre nodos.
- Ciclo de vida de un programa Spark
- Configuración
- Ejecución de tareas: spark-submit (dependencias, etc), REPL (spark-shell, pyspark), notebooks
- Acceso a datos: E/S (local, HDFS, etc).
- Interfaces de monitorización y análisis: Spark UI, Spark Master, Spark History Server
- Virtualización de infraestructura
 - Infraestructura local vs Cloud
 - Infraestructura como Servicio (IaaS)
 - Cloud privado: Propuestas Openstack y OpenNebula
 - Cloud público: Propuestas de IBM Softlayer, Amazon EC2, Rackspace, Google Cloud y Microsoft Azure
 - Cloud público vs Cloud privado
 - Prácticas: Despliegues Cloud privado
 - Prácticas: Monitorizando en el Cloud público
 - Prácticas: Comparativa de rendimiento en el Cloud
- Virtualización basada en contenedores
 - Diseño de aplicaciones en contenedores
 - Gestión de imágenes y versiones
 - Orquestación y Comunicación
 - Seguridad
- Plataformas como servicio (PaaS): Bluemix
 - Concepto de Plataforma como Servicio
 - Utilidades y nuevos modelos de consumo de servicios
 - La plataforma Bluemix
 - Ejercicios prácticos con Bluemix.

BIBLIOGRAFÍA

1. (A completar por los profesores de la asignatura)
- 2.

MÉTODOS DOCENTES

- Lección magistral
- Resolución de problemas
- Prácticas de laboratorio

- Estudio de casos

MÉTODOS DE EVALUACIÓN

- Asistencia a clase: **10%**
- Evaluación continua: **40%**
- Examen final: **50%**