

# Sesión de Laboratorio de Modelos Avanzados de Bases de Datos

Subject: **Hadoop-HDFS-HBase**

**Objetivo:** Crear una infraestructura capaz de recoger grandes cantidades de información generada mediante “Twitter” y una arquitectura Lambda<sup>1</sup> para su posterior análisis.

**Condiciones:** Se debe diseñar una base de datos capaz de recoger información mediante el API de Twitter de una temática determinada, con diferentes criterios, y almacenándola en el sistema HBASE para un posterior análisis de esta. El enunciado se divide en dos grandes partes:

1. Diseño del sistema e implementación y
2. Test de rendimiento, pruebas y optimización.

**Tecnologías a utilizar:**

- Máquinas virtuales para montar el sistema: ejemplo UBUNTU.
- Base de datos: HBase (Obligatoria).
- Análisis de resultado, a elegir por el alumno.

**Entrega:**

Este enunciado contiene 2 prácticas evaluables por separado:

**P03:** Modelado de la solución y preparación

- Memoria explicativa (15 hojas como máximo) donde se mostrará el modelo de datos, cómo se van a analizar posteriormente esos datos (Map&Reduce), y se desarrollará un cliente tipo.
- **Fecha límite de entrega:** 30/4/2020.

**P04:** Ejecución de múltiples clientes y múltiples consultas de análisis: resultados y acciones correctoras.

- Memoria explicativa (15 hojas máximo) y/o vídeo demostrativo del trabajo realizado.
- Código fuente de la solución aportada.
- **Fecha límite de entrega:** 17/5/2020.

---

<sup>1</sup> Sistema tolerante a fallos, tanto humanos como de hardware, que fuera linealmente escalable y que permitiese realizar escrituras y lecturas con baja latencia. Se caracteriza por utilizar distintas capas para el procesamiento batch y el streaming.