

# Ingeniería para el Procesado Masivo de Datos

Curso 2025/2026

**Dr. Pablo J. Villacorta**

# Objetivos de la asignatura

- ▶ Comprender la necesidad de tecnologías Big Data como Hadoop y Spark en la sociedad actual
- ▶ Conocer el stack tecnológico de Hadoop y algunos componentes fundamentales: HDFS, Spark, Hive, Kafka, Impala
- ▶ Entender para qué sirven, cómo se relacionan y sus fundamentos de uso, y ponerlos en práctica
- ▶ Conocer los fundamentos de las plataformas Cloud

# Temario

- ▶ Tema 1: Introducción a las Tecnologías Big Data
- ▶ Tema 2: HDFS (y un vistazo a MapReduce)
- ▶ Tema 3: Spark I (Arquitectura y RDDs)
- ▶ Tema 4: Spark II (Spark SQL)
- ▶ Tema 5: Spark III (MLlib, Structured Streaming)
- ▶ Tema 6: Apache Kafka
- ▶ Tema 7: Hive e Impala
- ▶ Tema 8: Cloud Computing I – Microsoft Azure
- ▶ Tema 9: Cloud Computing II – Amazon Web Services
- ▶ Tema 10: Cloud Computing III – Google Cloud Platform

Lunes 12 de mayo a  
las 23:59

Actividad 1  
grupal (5.0)

Lunes 9 de junio a  
las 23:59

Actividad 2  
individual (3.8)

Domingo 29 de junio  
a las 23:59

Actividad 3  
grupal (5.0)

# Herramientas

- ▶ Para Hadoop: entorno en Google Cloud Platform (GCP)
  - ▶ Clúster desplegado con Dataproc (JupyterLab + Spark + HDFS + Kafka)
- ▶ Objetivos para la parte de Hadoop: entender la esencia de cada herramienta (**para qué** se usa) y verla en pequeños casos prácticos
- ▶ Objetivos para la parte de Cloud Computing: entender los fundamentos, similitudes y diferencias entre plataformas cloud

# Evaluación

- ▶ Evaluación continua: 40 % de la calificación final
  - ▶ Hasta 15 puntos disponibles, que saturan en 10
  - ▶ 13.8 puntos son actividades prácticas (la primera en laboratorio)
  - ▶ 0.2 puntos por asistencia a al menos 2 sesiones presenciales
  - ▶ 1 puntos de tests de autoevaluación
- ▶ Examen presencial: 60 % de la calificación final
  - ▶ Imprescindible aprobarlo (5/10) para hacer media
  - ▶ Formato todavía por determinar
  - ▶ Podría incluir preguntas de respuesta múltiple

# Sobre las actividades evaluables

- ▶ Se corregirá tal cual el fichero que hayáis entregado, **sin excusas**.
- ▶ *No se permitirán entregas posteriores por "equivocación al subir el fichero".*
  - ▶ Notebook en blanco = 0 en esa actividad
  - ▶ Si es grupal, **todos los alumnos deben subir los mismos ficheros**
  - ▶ **La calificación es individual**: si un miembro entrega un fichero con errores, diferente al de otros miembros, se le calificará por su fichero.
  - ▶ Se aconseja no apurar el plazo de entrega (las prisas llevan a esta clase de errores)
- ▶ Las fechas de entrega de las actividades son **definitivas**.
  - ▶ No se va a ampliar el plazo de entrega cuando se acerque la fecha
- ▶ Tenéis desde hoy el calendario completo de entregas de todas las asignaturas  
→ debéis distribuir vuestro tiempo
  - ▶ Coincidirán entregas de varias asignaturas en la misma semana y/o en el mismo día. La fecha de entrega es solo "tope", pero podéis entregar antes.

# Dudas, sugerencias y comentarios

- ▶ Durante las clases o a través de la plataforma
- ▶ **Nunca** dudéis en preguntar
  - ▶ Casi siempre habrá otras personas con la misma duda o muy similares, y la respuesta le servirá a toda la clase
  - ▶ Una duda sin aclarar en algo simple puede llevar a conclusiones erróneas más adelante
  - ▶ Para evitarlo: preguntar siempre, cualquier cosa, en cualquier momento
  - ▶ No se responderán preguntas sobre "cómo será el examen" (para esto tenemos la última sesión de la asignatura)

...y ahora, vamos a empezar  
:-)

Cualquier duda, consulta o comentario:  
mensaje a través de la plataforma!

**unir**  
LA UNIVERSIDAD  
EN INTERNET

[www.unir.net](http://www.unir.net)