

# Grado en Ingeniería Informática, curso 2024-25

## Computación en la nube

### Manual de actividad practica 7: Subida de datos

El objetivo de esta práctica es aprender a gestionar la ingesta y procesamiento de datos a gran escala en AWS, utilizando servicios como S3, Kinesis, y AWS Glue.

#### Actividades

1. Configurar un bucket S3 con la estructura de carpetas adecuada para almacenar:
  - Datos sin procesar (raw)
  - Datos procesados (processed)
  - Scripts y configuraciones (config)
2. Implementar un productor de datos que genere registros simulados (logs, métricas o transacciones) y los envíe a un Kinesis Data Stream
3. Configurar un consumidor usando Kinesis Firehose que:
  - Recoja los datos del stream
  - Realice una transformación básica (p.ej., filtrado o enriquecimiento)
  - Almacene los resultados en el bucket S3 en la carpeta de datos procesados
4. Utilizar AWS Glue para:
  - Crear un crawler que analice los datos en S3
  - Definir un esquema en el Catálogo de Datos
  - Crear y ejecutar un trabajo ETL simple
5. Actividad extra (Opcional): Implementar un panel de control en Quick-Sight que visualice los datos procesados

**Evaluación** Esta práctica se evaluará sobre 10 puntos repartidos de la siguiente forma:

1. Defensa de actividades propuestas (5 puntos): Se realizarán las actividades propuestas y se evaluará haber cumplido los objetivos en una defensa con el profesor en el aula.
2. Entrega de memoria (3 puntos): Se evaluará la memoria de la práctica entregada en el moodle en formato PDF. Se tendrá en cuenta la presentación y el contenido de la misma. Esta deberá incluir si corresponde:
  1. Descripción de los pasos seguidos para cumplir los objetivos
  2. Código del productor de datos y configuraciones de Kinesis
  3. Diagrama del flujo de datos implementado

4. Presupuesto y estimación de gasto de los recursos desplegados
3. Actividades extras (2 puntos): El alumno podrá proponer una actividad extra a negociar con el profesor.