

# Estrategias y Herramientas de Computación Big Data en la Nube

## Trabajos de la Asignatura

Germán Moltó

Departamento de Sistemas Informáticos y  
Computación - Instituto de Instrumentación para  
Imagen Molecular

[gmolto@dsic.upv.es](mailto:gmolto@dsic.upv.es)

<http://www.grycap.upv.es/gmolto>

Master Universitario en Computación en la Nube  
y de Altas Prestaciones

# Resultados de Aprendizaje

- Se espera que tras esta presentación seas capaz de:
  - Recordar la importancia de la realización del trabajo de asignatura.
  - Conocer el listado de trabajos ofertado por el profesor.
  - Decidir qué trabajo elegir o realizar una propuesta de trabajo al profesor.

# Características del Trabajo

- El trabajo final de asignatura:
  - Contribuye un 40% a la calificación final, de acuerdo a la normativa de la asignatura.
  - Debe integrar varios de los servicios tratados en la asignatura para resolver un problema con una cierta complejidad.
- Se subirá la siguiente documentación por tareas de PoliformaT antes de la fecha límite :
  - una breve memoria que resumirá el trabajo realizado.
  - el código desarrollado, si aplica.
- Manda un correo electrónico a tu profesor en cuanto hayas decidido el trabajo que quieres hacer.
  - *First Come, First Served*

# Trabajo 1: Análisis de Datos con MapReduce

- Objetivo
  - Implementar un código de análisis de datos usando el modelo de programación MapReduce sobre Apache Hadoop.
  - Conjunto de datos a elección del alumno (autorizado por el profesor).
- Tecnologías involucradas
  - Modelo de Programación MapReduce
  - Framework Apache Hadoop
  - Implementación mediante Hadoop Streaming

# Ejemplos de DataSets

- Precio pagado por propiedades de Inglaterra y Gales.
  - <https://www.gov.uk/government/collections/price-paid-data>
- Multas de circulación
  - <http://datos.gob.es/es/catalogo/I01280796-multas-de-circulacion>
- Reseñas de videojuegos vendidos en Amazon
  - <http://snap.stanford.edu/data/web-Amazon-links.html>
- Datos de movilidad de estudiantes Erasmus
  - <https://data.europa.eu/euodp/es/data/dataset/erasmus-mobility-statistics-2013-14/resource/1827f69e-ec9e-47ac-b67b-04062429d955>
- Repositorios de datasets
  - <https://data.gov.uk/>
  - <http://datos.gob.es>
  - <http://data.europa.eu>
- Netflix: <https://www.kaggle.com/ashishgup/netflix-rotten-tomatoes-metacritic-imdb>

# Trabajo 2: Computación Basada en la Nube

- Objetivo
  - Utilizar cualquier tecnología/servicio que involucre el uso de computación intensiva en la nube
  - Ejemplos:
    - Alces Flight: <https://alces-flight.com>
    - Kubernetes Jobs
      - <https://kubernetes.io/docs/concepts/workloads/controllers/job-run-to-completion/>
    - AWS Fargate: <https://aws.amazon.com/es/fargate/>
    - AWS Batch: <https://aws.amazon.com/es/batch/>
- Implica definir un caso de uso (sencillo) y resolverlo con alguna de estas tecnologías / servicios.

## Trabajo 3: Tú Decides

- Si no te gusta ninguno de los trabajos propuestos o tienes alguna idea en mente también puedes plantearla como trabajo.
- Asegúrate de comentar la idea al profesor para que te autorice la realización del trabajo en el contexto de esta asignatura.