

Estrategias y Herramientas de Computación Big Data en la Nube

Presentación de la Asignatura Máster Universitario en Computación en la Nube y de Altas Prestaciones

Germán Moltó

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

gmolto@dsic.upv.es

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

Resultados de Aprendizaje

- Se espera que tras esta presentación seas capaz de:
 - Conocer al profesor de la asignatura.
 - Conocer el contexto y alcance de esta asignatura.
 - Conocer la organización y estructuración de la asignatura.
 - Comprender el mecanismo de evaluación de la asignatura.

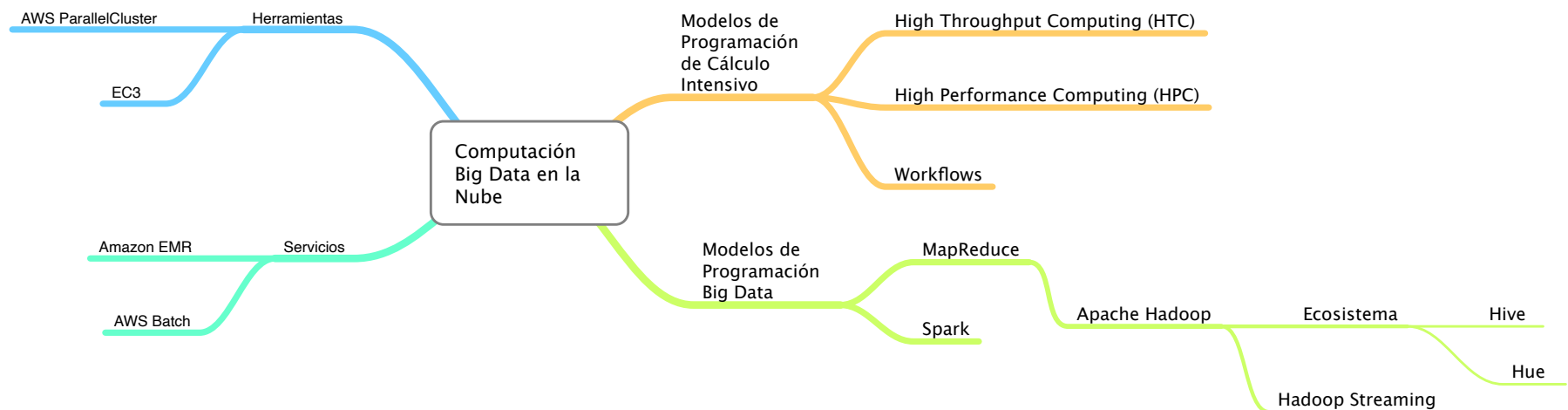
Profesor de la Asignatura

- Germán Moltó – gmolto@dsic.upv.es
 - Catedrático de Universidad en el Departamento de Sistemas Informáticos y Computación
 - Investigador en el Instituto de Instrumentación para Imagen Molecular
 - Área de Grid y Cloud de Altas Prestaciones
- Imparte asignaturas de Cloud en:
 - Máster en Computación en la Nube y de Altas Prestaciones
 - Máster en Big Data Analytics
 - Máster en Ciberseguridad y Ciberinteligencia
 - Curso Online de Cloud Computing con Amazon Web Services
 - Grado en Ciencia de Datos
 - Grado en Ingeniería Informática



Estrategias y Herramientas de Computación Big Data en la Nube

- Modelos de programación de cálculo intensivo, orientados a Big Data y computación intensiva.



Resultados de Aprendizaje (de la Asignatura)

- Entender los principales modelos de programación orientados a cálculo intensivo y su aplicación en la nube.
- Realizar el despliegue de clusters virtuales elásticos en la nube para soportar aplicaciones HPC y HTC.
- Utilizar el modelo MapReduce y su implementación mediante Hadoop para resolver problemas de mediana complejidad.
- Conocer Spark y utilizarlo para resolver problemas de mediana complejidad.
- Usar Hive para consultar conjuntos de datos sobre clusters Hadoop.

Temario de la Asignatura

- Tema 1. Computación Intensiva en la Nube
- Tema 2. Procesado de Big Data en la Nube

Enfoque de la Asignatura

- Este seminario es eminentemente práctico.
- Se realizan prácticas con:
 - Amazon Web Services (AWS), el proveedor Cloud líder en el mercado.
 - Amazon EMR (Elastic MapReduce)
 - AWS ParallelCluster, una herramienta para el despliegue de clusters virtuales sobre Amazon EC2.
 - Apache Hadoop, una implementación del modelo de programación MapReduce.
 - Apache Hive, una herramienta de data warehouse para consultar grandes conjuntos de datos.
 - Apache Hue, una interfaz gráfica para facilitar la interacción con un cluster Hadoop.
 - También veremos conceptos básicos sobre Apache Spark.

Evaluación de la Asignatura

- Evaluación de CBD:
 - 20% de evaluación de portafolio (incluye ejercicios propuestos a lo largo de las prácticas).
 - 40% calificación de un test de preguntas cerradas (a realizar sobre PoliformaT), indicado como prueba escrita en la Guía Docente
 - 40% de realización de un trabajo académico a elegir de una lista de trabajos propuestos o planteado por el alumno con el visto bueno del profesor.
 - Ver detalles en la tarea de PoliformaT

Sobre el Cluster Hadoop

- Accesible mediante SSH con las credenciales de usuario recibidas en el correo de bienvenida (mismas que para acceso a la instancia de prácticas de AWS)
- Importante:
 - 8 GB de RAM en el nodo principal
 - Solo 8 GB en la partición /
 - 50 GB en la partición /mnt/disk
- Consejos
 - No copies ficheros grandes (> 200 MB) en tu carpeta de usuario
 - Clúster virtualizado
 - Haz copia de seguridad de los ficheros de las actividades (*scp* o *copy/paste* del terminal)
 - Se recomienda realizar los trabajos en grupos de 3 personas

Sobre el Portafolio

- Representa un 20% de la calificación e incluye la resolución de todas las actividades de las prácticas y de las actividades propuestas evaluables durante las prácticas:
 - Práctica de MapReduce
 - Actividades de Spark
- Entregar un pequeño resumen con las aproximaciones propuestas (incluyendo código si es necesario) mediante tareas de PoliformaT, antes de una semana tras la última sesión (ver detalles en la tarea de PoliformaT).

¿Dudas o Cuestiones?

Germán Moltó

Despacho 1D19 – DSIC

Universitat Politècnica de València

gmolto@dsic.upv.es

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

Tutorías bajo demanda