

Procesamiento de Datos Masivos

Tutoría Inicial



Universidad
Internacional
de Valencia

Máster en Big Data y Ciencia de Datos

03MBID

Procesamiento Masivo de Datos

Profesores:

Yudith Cardinale/Irvin Dongo (Grupo A)

Òscar Garibo Orts (Grupo B)



Los profesores

Los alumnos

La asignatura

Motivación

Objetivos

Contenido

Planificación de sesiones

Actividades del portafolio

Fechas de las evaluaciones



viu

Universidad
Internacional
de Valencia

Contenido

Los profesores

Los alumnos

La asignatura

Motivación

Objetivos

Contenido

Planificación de sesiones

Actividades del portafolio

Fechas de las evaluaciones

Presentación de los profesores

Yudith Cardinale



Universidad
Internacional
de Valencia

- ▶ **Asesoría técnica en instituciones gubernamentales en Venezuela:** PDVSA, CNTI, CNE, Vicepresidencia, CENDIT, CANTV, Alcaldía Libertador, Hospital pediátrico infantil

- ▶ **Profesora/Investigadora**

- ▶ Universidad Internacional de Valencia, España
- ▶ Universidad Simón Bolívar, Venezuela
- ▶ Universidad Central de Venezuela
- ▶ Universidad Católica Andrés Bello, Venezuela
- ▶ Universidad Centroccidental Lisandro Alvarado, Venezuela
- ▶ Universidad de Los Andes, Venezuela
- ▶ Universidad Católica San Pablo, Perú
- ▶ Universidad Nacional de Ingeniería, Perú

- ▶ **Investigadora**

- ▶ Université de Pau, Biarritz, Francia
- ▶ Université Paris 13, Francia
- ▶ Université de Nanterre, Francia
- ▶ Université Paris Dauphine, Francia
- ▶ Universidad de Valparaíso, Chile



Presentación de los profesores

Yudith Cardinale

Proyectos de investigación:

- ▶ **RUTAS:** Robots para centros Urbanos Turísticos, Autónomos y basados en Semántica
- ▶ **PASEO:** Profile generation And content Suggestion on E-tOurism
- ▶ **T-CREO:** Twitter **CRE**ditability Analysis FramewOrk
- ▶ High Performance Computing
- ▶ Web Semántica, ontologías
- ▶ Casas inteligentes, IoT, *Healthcare*
- ▶ GreenSE4IoT: Towards Energy-efficient Software for Distributed Systems ... y otros temas más
- ▶ **GRID:** **GR**upo de **I**ntestigación en ciencia de **D**atos de la **VIU**

Entretención

- ▶ Fútbol
- ▶ Montañismo, senderismo, ...
- ▶ Carrera ... hasta de 21 km
- ▶ Cine, lectura, juegos de mesa
- ▶ Viajar ... [Venezuela](#), [España](#), [Francia](#), [Alemania](#), [Panamá](#), [Perú](#), [Chile](#), [USA](#) ...



Presentación de los profesores

Irvin Dongo



► Profesor/Investigador

- Universidad Internacional de Valencia, España
- Universidad Simón Bolívar, Venezuela
- Universidad Católica San Pablo, Perú

► Investigador

- ESTIA, Francia
- Universidad de Valparaíso, Chile



Presentación de los profesores

Irvin Dongo



Universidad
Internacional
de Valencia

► Proyectos de investigación:

- **RUTAS**: Robots para centros Urbanos Turísticos, Autónomos y basados en Semántica
- **T-CREO**: Twitter **CR**edibility Analysis Framew**Ork**
- Web Semántica, ontologías
- **GRID**: **GR**upo de Investigación en ciencia de Datos de la VIU

► Entretenimiento

- Ajedrez
- Juegos de mesa
- Cine



Presentación de los profesores

Òscar Garibo Orts



Universidad
Internacional
de Valencia

► Experiencia en la empresa:

- Consultoría TIC con proyectos de almacenamiento y distribución de datos
- Soluciones para Smart City
- Formación TIC

► Profesor/Investigador

- Universidad Internacional de Valencia, España
- Universitat Politècnica de València

► Investigador

- Universidad Internacional de Valencia, España
- Universitat Politècnica de València

Presentación de los profesores

Òscar Garibo Orts

► Proyectos de investigación:

- **VRAIN**, Valencian Research Institute for Artificial Intelligence
- **IUMPA**, Institut Universitari de Matemàtica Pura i Aplicada
- **GRID**: GRupo de Investigación en ciencia de Datos de la VIU ... y alguna que otra cosa

► Entretenimiento

- Baloncesto
- Senderismo
- Cine, teatro, lectura
- Juegos de mesa



Contenido

Los profesores

Los alumnos

La asignatura

Motivación

Objetivos

Contenido

Planificación de sesiones

Actividades del portafolio

Fechas de las evaluaciones

Presentación de los alumnos

- ▶ Perfil profesional
- ▶ Experiencia técnica previa
- ▶ Interés en el máster
- ▶ Interés en la asignatura
- ▶ ¿Linux o Windows?
- ▶ Película favorita/Libro favorito
- ▶ Equipo de fútbol favorito ... o mejor no



Contenido

Los profesores

Los alumnos

La asignatura

Motivación

Objetivos

Contenido

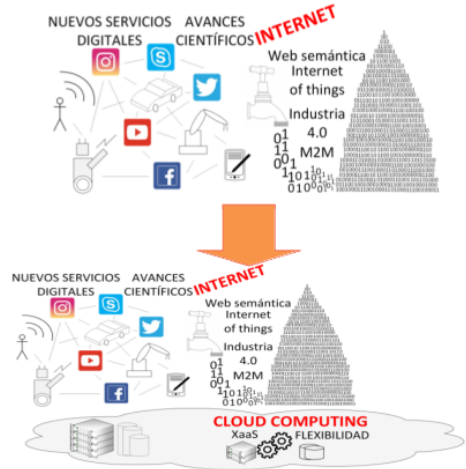
Planificación de sesiones

Actividades del portafolio

Fechas de las evaluaciones

Escenario de motivación

- ▶ Enorme cantidad de información accesible a través de Internet
⇒ Necesidad de procesar eficientemente grandes cantidades de datos
- ▶ Capacidad de almacenamiento y procesamiento local insuficiente
⇒ Desarrollo de nuevas técnicas de almacenamiento y procesamiento en sistemas paralelos y distribuidos en la nube (*Cloud Computing*)



La asignatura

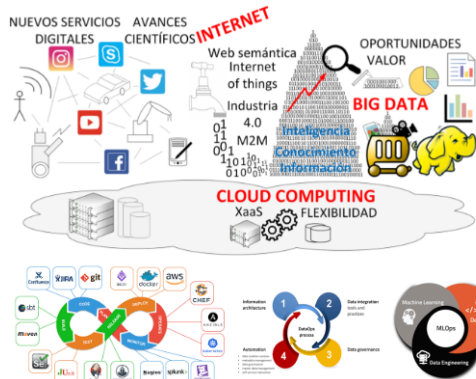
Motivación



Universidad
Internacional
de Valencia

Escenario de motivación (cont.)

- ▶ El *cloud computing* permite el acceso transparente, seguro y barato a enormes recursos computacionales para almacenamiento de datos, procesamiento y análisis, despliegue de aplicaciones ...
- ▶ Este nuevo reto en el procesamiento masivo de datos, también demanda
 - ⇒ nuevas metodologías, guías y buenas prácticas en el desarrollo de aplicaciones y servicios de gestión y procesamiento de datos
 - ⇒ herramientas de automatización de despliegue, monitoreo, trazabilidad, gestión de la configuración, ...



La asignatura

Objetivos

- ▶ La asignatura busca familiarizar al estudiante con diferentes técnicas de **procesamiento de grandes cantidades de información**.
- ▶ Conocer los principales modelos del *Cloud Computing*, modelos de servicios y modelos de despliegue.
- ▶ Desplegar un *cluster virtual* para el procesamiento del Big Data usando Hadoop.
- ▶ Conocer el **modelo de programación MapReduce** e implementar algoritmos para resolver problemas simples en los principales frameworks: **Hadoop MapReduce**, **Spark** y **Flink**.
- ▶ Utilizar herramientas y lenguajes de alto nivel para el procesamiento masivo de datos en **Hadoop** y **Spark**.
- ▶ Utilizar **Apache Spark** como motor de análisis unificado para el procesamiento de datos masivos, combinando SQL, procesamiento de flujos y analítica compleja.



Universitat
Internacional
de València



La asignatura

Contenido



Universidad
Internacional
de Valencia

- ▶ Tema 1: Introducción al Cloud Computing
- ▶ Tema 2: Introducción al Big Data
- ▶ Tema 3: Ecosistema Hadoop: Procesamiento distribuido
- ▶ Tema 4: Ecosistema Spark: Procesamiento en memoria
- ▶ Tema 5: Ecosistema Flink: Procesamiento streaming



La asignatura

Planificación de sesiones

Sesión	Fecha	Contenido/Tema
SESIÓN 1	17/05/2023	Tutoría colectiva. Tema 1: Introducción al <i>Cloud Computing</i> Prof. Garibo
SESIÓN 2	22/05/2023	Tema 1: Introducción al <i>Cloud Computing</i> Tema 2: Introducción al Big Data Prof. Garibo
SESIÓN 3	24/05/2023	Tema 2: Introducción al Big Data Tema 3: Ecosistema Hadoop Prof. Garibo
SESIÓN 4	29/05/2023	Tema 3: Ecosistema Hadoop Actividad guiada Prof. Garibo
SESIÓN 5	31/05/2023	Tema 3: Ecosistema Hadoop Actividad guiada Prof. Garibo
SESIÓN 6	05/06/2023	Tema 3: Ecosistema Hadoop Actividad guiada Prof. Garibo
SESIÓN 7	07/06/2023	Tema 4: Ecosistema Spark Prof. Cardinale/Prof. Dongo
SESIÓN 8	12/06/2023	Tema 4: Ecosistema Spark Actividad guiada Prof. Cardinale/Prof. Dongo
SESIÓN 9	14/06/2023	Tema 4: Ecosistema Spark Actividad guiada Prof. Cardinale/Prof. Dongo
SESIÓN 10	19/06/2023	Tema 4: Ecosistema Spark Actividad guiada Prof. Cardinale/Prof. Dongo
SESIÓN 11	21/06/2023	Tema 5: Ecosistema Flink Prof. Cardinale/Prof. Dongo
SESIÓN 12	26/06/2023	Resumen Final: Ecosistema Big Data. Debate en grupos. Tutoría colectiva. Prof. Cardinale/Prof. Dongo
SESIÓN 13	28/06/2023	Seminario de experto

NOTA: El Grupo A tiene horario de 18H a 20H y el Grupo B de 20H a 22H

La asignatura

Actividades del portafolio



Universitat
Internacional
de València

- ▶ **Actividad 1 (20%)**: Implementar programas Big Data utilizando el framework Apache MapReduce (**grupo de 2 estudiantes**).
 - ▶ **Actividad 2 (20%)**: Implementar programas Big Data utilizando el framework Apache Spark (**individual**).
 - ▶ **Foro debate (10%)**: Participar en el sesión de debate en grupo sobre un tema relacionado con la asignatura (**grupo de 2 estudiantes**).
 - ▶ **Actividad Evaluación Continua (10%)**: Test autoevaluación sobre vídeo relacionado a un tema de la asignatura (**individual**).
- ¡Atención: el foro debate y el test autoevaluación sólo están disponibles para primera convocatoria!**



La asignatura

Actividades del portafolio

- ▶ Para la Actividad 1 y Actividad 2 deben elaborar un informe en pdf mostrando los screenshots (pantallazos) de los pasos ejecutados para realizar la actividad, acompañado de explicaciones claras de lo realizado.
- ▶ La tarea se entregará en algún formato comprimido (gzip, zip, etc.). El fichero comprimido debe contener tanto documentos en PDF como códigos en los lenguajes de programación solicitados, así como instrucciones para la compilación y ejecución de dichos códigos estándares, entregarlo a través del apartado correspondiente a Actividades en la página de la asignatura en el Campus Virtual y tendrá el siguiente formato:
AG_X-02MBIG-Apellido-Nombre.gz.
- ▶ La participación en el Foro Debate la realizarán a través del apartado de Foros en la página de la asignatura en el Campus Virtual.
- ▶ **¡Atención: El foro debate y el test de autoevaluación sólo están disponibles para primera convocatoria!**



La asignatura

Fechas de las evaluaciones: Portafolio (60%)

Fechas de entrega del portafolio	
1ª Convocatoria	<p>Actividad 1: Recomendada: Miércoles 07 de junio de 2023 hasta las 23:59 Fecha última: Lunes 03 de julio de 2023 hasta las 23:59</p> <p>Actividad 2: Recomendada: Viernes 30 de junio de 2023 hasta las 23:59 Fecha última: Lunes 03 de julio de 2023 hasta las 23:59</p> <p>Foro debate: Recomendada: Lunes 12 de junio de 2023 hasta las 23:59 Fecha última: Lunes 03 de julio de 2023 hasta las 23:59 Solamente disponible en 1era convocatoria</p> <p>Test autoevaluación (vídeo): Recomendada: Miércoles 14 de junio de 2021 hasta las 23:59 Fecha última: Lunes 03 de julio de 2023 hasta las 23:59 Solamente disponible en 1era convocatoria</p>
2ª Convocatoria	<p>Actividad 1, Actividad 2 (no hay test ni foro debate): Miércoles 25 de Octubre de 2023 hasta las 23:59</p>



¡Atención!: Luego de la fecha de entrega de cada actividad, pierden el 20% por entrega tardía.

La asignatura

Fechas de las evaluaciones: Prueba final (40%)

Fechas de realización de la prueba		
Franja	A	B
1ª Convocatoria	Lunes 3 de julio de 2023 de 12:00 a 14:00	Lunes 3 de julio de 2023 de 118:00 a 20:00
2ª Convocatoria	Miércoles 25 de Octubre de 2023 de 12:00 a 14:00	Miércoles 25 de Octubre de 2023 de 18:00 a 20:00



- ▶ El examen constará de alrededor de 20 preguntas de tipo test, junto con 5 a 10 preguntas de respuestas cortas (en total entre 25 y 30 preguntas). El tiempo establecido para el examen será de 90 minutos.
- ▶ **¡Atención!: Deben aprobar tanto el Portafolio como la prueba final para aprobar la asignatura. Adicionalmente, deben aprobar la Actividad 1 y la Actividad 2.**

¡Mucho éxito!



Universidad
Internacional
de Valencia

Contactos:

yudith.cardinale@campusviu.es

oscar.garibo@campusviu.es

irvin.dong@campusviu.es

