Gestión de *big data*: tecnologías

Francesc Julbe
Mario Macias Lloret
Jordi Nin

PID_00237565

Material docente de la UOC

UOC
Universitat Oberta de Catalunya

Francesc Julbe

Ingeniero de Telecomunicaciones por la UPC (2001). En sus inicios trabajó en empresas del sector TIC y de consultoría tecnológica como Agilent Technologies y Grupo Penteo realizando tareas de desarrollo de software y siendo responsable de innovación hasta 2004, cuando se incorporó al proyecto Gaia con el equipo de ingeniería de la UB. En 2007 se incorporó al European Southern Observatory (ESO) en Munich como ingeniero sénior en diferentes proyectos de desarrollo relacionados con la gestión de los archivos astronómicos del Very Large Telescope (VLT) y Atacama Large Milimiter Array (ALMA). En 2012 se reincorporó al proyecto Gaia en el equipo de ingeniería de Barcelona como *GOG Project Manager*. Desde 2014 lidera el desarrollo de una plataforma de *Big Data* para la misión Gaia, tanto desde un punto de vista de arquitectura como de casos de uso científico que se implementan sobre ella.

Mario Macias Lloret

Ingeniero de software de sistemas en Midokura S.L., trabajo que compagina con el de profesor a tiempo parcial en la Universitat Politècnica de Catalunya. Durante los últimos 10 años ha trabajado en proyectos de investigación sobre *cloud*, *big data* y *green computing* en el Barcelona Supercomputing Center. Actualmente trabaja en Midokura, empresa dedicada a las soluciones de redes virtuales para centros de datos e Internet de las Cosas.

Jordi Nin

Máster y doctorado en Ciencias de la Computación por la Universidad Autónoma de Barcelona (UAB, 2005, 2008, respectivamente). Obtuvo el premio a la mejor tesis del año en el departamento de informática de la UAB. De 2006 a 2008, fue becario predoctoral en el IIIA-CSIC. Posteriormente, se trasladó durante dos años como investigador postdoctoral Marie Curie en el LAAS-CNRS (Toulouse, Francia). De 2010 a 2015, trabajó como profesor lector en el departamento de arquitectura de computadores de la UPC y como investigador asociado en el Barcelona Supercomputing Center (BSC). Hoy en día, trabaja como Senior Data Scientist en BBVA Data & Analytics. Sus intereses de investigación se centran en el aprendizaje automático y análisis de grandes volúmenes de datos. Durante su carrera, ha realizado más de 80 publicaciones en revistas científicas, conferencias o capítulos de libros. Además, ha estado involucrado en varios proyectos de investigación, tanto nacionales como internacionales.

El encargo y la creación de este material docente han sido coordinados por el profesor: Jordi Casas Roma (2016)

Primera edición: septiembre 2016
© Francesc Julbe, Jordi Nin, Mario Macias Lloret
Todos los derechos reservados
© de esta edición, FUOC, 2016
Av. Tibidabo, 39-43, 08035 Barcelona
Diseño: Manel Andreu
Realización editorial: Oberta UOC Publishing, SL
Depósito legal: B-16.654-2017

Ninguna parte de esta publicación, incluido el diseño general y la cubierta, puede ser copiada, reproducida, almacenada o transmitida de ninguna forma, ni por ningún medio, sea éste eléctrico, químico, mecánico, óptico, grabación, fotocopia, o cualquier otro, sin la previa autorización escrita de los titulares del copyright.

Contenidos

Módulo didáctico 1
Identificación, captura y almacenamiento de datos masivos
Jordi Nin

1. Qué entendemos por datos masivos o *big data*
2. Estructura general de un sistema de *big data*
3. Sistema de archivos
4. Sistema de cálculo distribuido
5. Gestor de recursos
6. Escenarios de procesamiento distribuido

Módulo didáctico 2
Big data frameworks
Francesc Julbe

1. Modelos de procesamiento distribuido en *big data*
2. Procesado distribuido en modo *batch*
3. Procesado en *streaming*
4. Otras herramientas *big data*
5. Soluciones Vendor

Módulo didáctico 3
Análisis y visualización de datos
Mario Macias Lloret

1. *Streaming*: análisis de datos en tiempo real
2. Aprendizaje autónomo
3. Análisis de grafos
4. Visualización de datos masivos
5. Otras herramientas para el análisis y la visualización de datos