

Ciclo de charlas GAMES Python para *Data Science*

consultas: ftobar@dim.uchile.cl

<http://games.cmm.uchile.cl/>

Charla 1. *Data Science*: Herramientas, lenguajes y Python

Gonzalo Ríos (Ingeniero CMM y estudiante de doctorado DIM)

10/6/2016, 14:30 hrs, Sala John von Neumann, CMM.

Abstract: *Data Science* es un área interdisciplinaria dedicada extraer conocimiento de grandes volúmenes de datos, utilizando diversas técnicas de *Machine Learning*, para resolver problemáticas de distinta naturaleza. La cantidad de datos y complejidad de los cálculos necesarios en aplicaciones de *Data Science* son en general muy altas, por lo que la eficiencia computacional es crucial. En esta charla revisaremos herramientas requeridas para trabajar con datos reales, principalmente sobre programación científica, lenguajes de programación y librerías de alto desempeño. Concluiremos la charla destacando las ventajas de Python en *Data Science*: su gran flexibilidad y claridad como lenguaje en sí, su sólida comunidad, documentación disponible y código libre.

Charla 2. Introducción a Python para computación científica

Javier Correa (Postdoc, Departamento de Ingeniería Eléctrica)

17/6/2016, 14:30 hrs, Sala John von Neumann, CMM.

Abstract: Python ha logrado posicionarse como un competidor real dentro de los lenguajes de computación científica. En esta charla se introducirán los conceptos y estructuras básicas de Python interactivamente usando un *IPython notebook*. Además, se presentarán las principales herramientas disponibles en Python que ayudan a resolver distintas tareas de un ámbito científico. Veremos también el ambiente **interactive IPython/Jupyter**, que permite ejecutar cuadernos interactivos; la librería numérica **NumPy**, que permite cargar, guardar y manejar eficientemente arreglos multidimensionales de datos; la librería de gráficos **matplotlib** y la librería científica **SciPy**, que proporciona funcionalidades como minimización y búsqueda de ceros en funciones, generación de números (pseudo-)aleatorios, entre otras. Finalmente, revisaremos brevemente otras librerías científicas disponibles como **pandas**, **scikit-learn**, **PyMC** y **SymPy**.

Charla 3. Creando una aplicación científica con Python

Nicolás Loira (Postdoc CMM)

24/6/2016, 14:30 hrs, Sala John von Neumann, CMM.

Abstract: En esta sesión crearemos desde cero un *script* en Python para resolver un problema de análisis de datos. Revisaremos desde el parseo de parámetros, la lectura y filtraje de los datos, el procesamiento de ellos, *profiling* para buscar problemas de performance y test unitarios, llegando hasta la creación de paquetes Python que pueden ser compartidos con la comunidad.

Charla 4. Python paralelo y distribuido

Juan Carlos Maureira (Scientist CMM)

1/7/2016, 14:30 hrs, Sala John von Neumann, CMM.

Abstract: En este tutorial mostraremos las capacidades existentes en Python para ejecutar código distribuido y paralelo. Cubriremos desde técnicas básicas como MP (*multiprocessing*) hasta SPARK (*big data*). En este camino, aprenderemos cómo aprovechar las ventajas de los dos RMS (*resource management systems*) más usados para HPC (*slurm* y *Sun Grid Engine*), como también como aplicar OpenMP a código C y Fortran. Finalizaremos este tutorial con una breve revisión de las optimizaciones necesarias para poder usar Python en un ambiente científico, desde su compilación hasta la inclusión de librerías optimizadas como MKL o cuBLAS.