

## **Curso Avanzado sobre Procesamiento de Grandes Volúmenes de Datos**

Mg. Robinson Rivas (Profesor Invitado). Profesor Universidad Central de Venezuela, Venezuela y  
Esteban Mocskos (Profesor Adjunto dedicación Exclusiva, FCEN-UBA)

### **Programa:**

Esta materia busca presentar herramientas prácticas para realizar procesamiento y análisis de datos a gran escala en el contexto de aplicaciones de inteligencia artificial y machine learning.

Se presentarán los fundamentos de Cloud Computing (Computación en la nube), las principales características, los diferentes modelos que han evolucionado desde la creación del concepto, y un repaso práctico por una experiencia de Computación en la Nube utilizando recursos reales en plataformas comerciales, a través de licencias gratuitas/educacionales.

Adicionalmente, se presentarán herramientas interactivas de exploración de grandes cantidades de datos provenientes de distintas aplicaciones científicas. Se darán los elementos para realizar visualizaciones in-situ. Se utilizarán herramientas basadas en código abierto como Jupyter o

JupyterLab, que permiten combinar el aspecto interactivo con cómputo reproducible con soporte a distintos escenarios científicos y tecnológicos. Al terminar este curso, los estudiantes

deben ser capaces de:

- Conocer la historia y los antecedentes de la computación en nube.
- Conocer qué es la computación en nube, a partir de ejemplos de problemas del mundo real.
- Comprender qué abarca un entorno en la nube.
- Comprender las ventajas de las ofertas en la nube en comparación con las instalaciones locales.
- Comprender los impactos que la nube tiene en los centros de datos tradicionales.
- Identificar y diferenciar varios componentes de infraestructura del centro de datos clásico y virtualizado.
- Explicar los requisitos de virtualización y las herramientas disponibles en cada capa de infraestructura de TI.
- Comprender los diversos modelos de servicio en la nube como IaaS, PaaS y SaaS.
- Comprender las ofertas de servicios básicos en la nube de Amazon Web Services (AWS), Microsoft Azure y/o Google Cloud.
- Conocer los aspectos comerciales de la computación en la nube.
- Comprender los diversos servicios en al menos un proveedor de servicios en la nube como Almacenamiento,
- Infraestructura, Computación, Datos, etc.
- Conocer lo requerido para construir servicios en la nube.
- Configurar una cuenta en al menos un proveedor de servicios en la nube.
- Realizar actividades prácticas sobre servicios de almacenamiento, cómputo, redes, y auto-administrados, en al menos un proveedor de servicios comerciales en la nube.

### **Temario**

- Introducción a la Computación en la Nube
  - Perspectiva histórica de la computación en nube.

- Estandarización, automatización, elasticidad rápida, fijación de precios.
- Arquitecturas y modelos de servicio en la nube (modelos de servicio y entrega, virtualización,
- hipervisores)
- Servicios en la Nube y Vendedores de IaaS, PaaS, SaaS (Amazon AWS, OpenStack, Google Cloud y
- Microsoft Azure, etc.)
- Anatomía de la nube (Capas en la nube, servicios web, etc.)
- Centros de Datos y Virtualización
  - Componentes de un Centro de Datos.
  - Abstracción de recursos.
  - Compartición de recursos.
  - Sandboxing.
- Servicios Básicos en la Nube
  - Introducción.
  - Servicios de Cómputo.
  - Servicios de Almacenamiento.
  - Servicios de Red.
- Bases de Datos en la Nube y software en la Nube
  - Introducción.
  - Bases de Datos Relacionales en la nube.
  - Bases de Datos No Relacionales en la nube.
  - Creación de arquitecturas escalables.
  - Confiabilidad, Desempeño, Seguridad,
- Optimización, Operaciones.
  - Microservicios.
  - Arquitectura de Aplicaciones Nativas para la Nube.
- Cómputo Interactivo
  - Introducción a Jupyter
  - Cómputo paralelo usando Jupyter
  - Visualización interactiva e in-situ.
  - Recorrido tecnológico desde ipywidgets a dashboards

## **Bibliografía**

- "Cloud Computing: Concepts, Technology & Architecture" by T. Erl, et al. (2013).
- "Cloud Native Infrastructure: Patterns for Scalable Infrastructure and Applications in a
- Dynamic Environment" by J. Fjordbak, et al. (2020).
- "Cloud Computing: From Beginning to End" by R. Buyya, et al. (2016).
- "Cloud Computing: Principles and Paradigms" by R. Buyya, et al. (2011).
- "Cloud Computing: A Practical Approach" by A. Sriram and S. Buyya (2013).
- "Cloud Computing and SOA Convergence in Your Enterprise: A Step-by-Step Guide" by D. Linthicum (2009).
- "Architecting the Cloud: Design Decisions for Cloud Computing Service Models" by M. Kavis (2014).
- "Enterprise Cloud Computing: Technology, Architecture, Applications" by G. Reese (2012).
- "Cloud Computing: SaaS, PaaS, IaaS, Virtualization, Business Models, Mobile, Security and More" by D. C. Marinescu (2011).

- "Mastering Cloud Computing" by R. Buyya, et al. (2013).
- "High Performance Computing with Python 3: How to Use Parallel Processing for Machine Learning" by Dr. Ronald K. Pearson (2019).
- "Parallel and High Performance Computing with Jupyter" by W. van Engen and J. J. van der Bijl (2020).
- "Jupyter Cookbook: Over 100 Hands-On Recipes to Quickly Prepare, Cook, and Serve Jupyter Notebooks" by Dan Toomey (2018).
- "High Performance Python: Practical Performant Programming for Humans" by Micha Gorelick and Ian Ozsvald (2020).
- "Data Science on the Google Cloud Platform: Implementing End-to-End Real-Time Data Pipelines: From Ingest to Machine Learning" by Valliappa Lakshmanan (2018).
- "High Performance Computing and Big Data Analytics: Concepts, Technologies, and Applications" by H. Zhang, et al. (2018).
- "Python for Data Science Handbook: Essential Tools for Working with Data" by Jake VanderPlas (2016).
- "Parallel Computing: On the Road to Exascale" by Gerhard R. Joubert, et al. (2015).
- "Python High Performance: Build High-Performance Concurrent, and Distributed Applications" by Gabriele Lanaro (2017).
- "Jupyter Notebooks: Quick Start Guide" by Packt Publishing (2018).