



**CLOUD COMPUTING.
APLICACIONES EN BIG DATA**

Año 2016

Carrera/ Plan:

Licenciatura en Informática Plan 2015

Licenciatura en Sistemas Plan 2015

*Licenciatura en Informática Plan 2003-07/
Plan 2012*

*Licenciatura en Sistemas Plan 2003-07/
Plan 2012*

Año: 4to

Régimen de Cursada: *Semestral*

Carácter: Optativa

Correlativas: Programación Concurrente

Coordinador: Ing. Armando De Giusti

Profesor: Lic. Ismael P. Rodríguez

Docente Auxiliar: Sebastián Rodríguez
Eguren

Hs. semanales: 6 hs.

FUNDAMENTACIÓN

Los avances en las tecnologías de virtualización y cómputo distribuido han dado origen al paradigma de Cloud Computing, que se presenta como una alternativa a los tradicionales sistemas de Clusters y Grids para ambientes de HPC y también como un esquema de procesamiento masivo de datos y de servicios remotos de gran impacto en la Informática actual.

Esta tecnología provee una abstracción de los recursos de hardware permitiendo ejecutar simultáneamente múltiples instancias de Software en máquinas virtuales sobre un único hardware físico (centralizado o distribuido).

Cloud Computing, proporciona grandes conjuntos de recursos físicos y lógicos (infraestructura, plataformas de desarrollo, almacenamiento y/o aplicaciones), fácilmente accesibles y utilizables por medio de una interfaz de administración web, con un modelo de arquitectura "virtualizada" Estos recursos son proporcionados como servicios ("as a service") y pueden ser dinámicamente reconfigurados para adaptarse a una carga de trabajo variable (escalabilidad), logrando una mejor utilización y evitando el sobre o sub dimensionamiento (elasticidad).

Al mismo tiempo uno de los campos de aplicación más importante para las arquitecturas Cloud es el de "big data" en el cual la capacidad de procesamiento se aplica al tratamiento masivo de información (numérica, textual, imágenes, señales, etc.), buscando extraer determinado conocimiento.



**UNIVERSIDAD NACIONAL DE LA PLATA
FACULTAD DE INFORMÁTICA**

Son objetivos de este curso: profundizar el conocimiento de la tecnología Cloud y realizar trabajo experimental en tratamiento masivo de datos, utilizando Cloud.

OBJETIVOS GENERALES:

- Profundizar el conocimiento de las arquitecturas CLOUD y su utilización.
- Analizar los problemas principales en las aplicaciones de Cloud Computing.
- Estudiar los fundamentos de los problemas de Big Data.
- Resolver problemas de Big Data, utilizando arquitecturas Cloud.

CONTENIDOS MINIMOS:

- Arquitecturas Cloud.
- Cloud Computing: Software de base y de aplicación.
- Estudio de clases de aplicaciones sobre Cloud. HPC.
- Conceptos de Big Data.
- Aplicaciones de big data sobre Cloud.

PROGRAMA ANALÍTICO

A. Fundamentos de Cloud Computing

- Evolución: de Clusters y GRID a Cloud.
- Análisis de beneficios del empleo de arquitecturas Cloud: Escalabilidad, Disponibilidad, Confiabilidad.
- Riesgos y desafíos del empleo de arquitecturas Cloud: Seguridad, Portabilidad, Eficiencia.
- Características de un Cloud.

B. Arquitecturas en Cloud Computing

- IaaS. Infraestructura como Servicio.
- PaaS. Plataforma como Servicio.
- SaaS. Software como Servicio.
- Tipos de arquitecturas Cloud. (públicas, privadas, híbridas).
- Virtualización estática y dinámica.
- Manejo de las Comunicaciones en Cloud.
- Manejo de Datos en Cloud.
- Seguridad en Cloud.



C. Software de Base y de aplicación en Cloud Computing

- Hipervisores.
- Balance de carga.
- Administración de recursos distribuidos.
- Programación de aplicaciones sobre Cloud.
- Configuración de máquinas virtuales.
- Métricas de performance.
- Eficiencia vs. Consumo.

D. Conceptos Fundamentales de Big Data

- Definición y dimensiones en Big Data.
- Introducción a las técnicas de almacenamiento. MapReduce.
- Aplicaciones de Big Data.
- Modelos de datos y modelos de procesamiento en Big Data

E. Aplicaciones de Big Data sobre Cloud.

- Algoritmos de tratamiento de información textual.
- Algoritmos de búsqueda inteligente.
- Algoritmos de tratamiento de imágenes.
- Estudio de casos.

METODOLOGÍA DE ENSEÑANZA

1) Modalidad presencial

La asignatura se estructura con clases teórico-prácticas y prácticas experimentales.

- Las clases teórico-prácticas son dictadas por los Profesores de la asignatura y son obligatorias para la promoción.
- Las explicaciones de práctica son introductorias al trabajo en Laboratorio, para facilitar la utilización del equipamiento y software por los alumnos. Se desarrollan en las clases teórico-prácticas.
- Se propone el desarrollo de trabajos concretos sobre configuraciones de Cloud, utilizando servicios disponibles sin costo (en el país y en el exterior). Estos trabajos abarcan aplicaciones de HPC y de servicios sobre Cloud.
- Se propone la resolución de problemas de Big Data, utilizando Cloud. Analizar las soluciones en función del rendimiento en tiempo y calidad de las respuestas. En principio con información textual e imágenes.
- Los trabajos se pueden realizar individualmente o en grupo de 2 personas. Las consultas y correcciones son realizadas en forma presencial o por medio de la plataforma de Educación a Distancia de la UNLP (WEBUNLP).

2) Modalidad no presencial

El alumno tendrá todo el material del curso a disposición por medio del entorno



**UNIVERSIDAD NACIONAL DE LA PLATA
FACULTAD DE INFORMÁTICA**

WebUNLP.

Se hace notar que para las tareas experimentales, el alumno deberá tener acceso a alguno de los Cloud habilitados por la cátedra, pero podrá hacerlo asincrónicamente.

Los alumnos deben aprobar las mismas entregas de los trabajos experimentales que los alumnos en modalidad presencial.

EVALUACIÓN

1) Modalidad presencial:

Los alumnos deben aprobar las entregas de los diferentes trabajos experimentales, para obtener la aprobación de los TP de la asignatura.

Posteriormente se les propondrá un trabajo final experimental, a defender en un coloquio, como Trabajo Final de Promoción, en una fecha de examen final.

2) Modalidad no presencial

Deben cumplir las entregas de los mismos trabajos experimentales para obtener la aprobación de los TP de la asignatura.

Deben rendir examen final, que puede incluir un trabajo final experimental, pero necesariamente contendrá preguntas de la Teoría presencial.

BIBLIOGRAFÍA BÁSICA

- Thomas Erl “**Cloud Computing. Concepts, Technology & Architecture**”. Prentice Hall 2013.
- Mayer-Schonberger, Cukier “**Big Data**”. Houghton Mifflin Harcourt Publishing 2012.
- Kirk, Hwu “**Programming massively parallel processors**”. Elsevier 2013.
- Grama A, Gupta A, Karypis G, Kumar V. “**Introduction to parallel computing**”. Second Edition. Pearson Addison Wesley, 2003.
- Furth, Escalante “**Handbook of Data Intensive Computing**”. Springer 2013.
- Rothon J. “**Cloud computing explained: Implementation handbook for enterprises**”. Recursive Press 2013.



BIBLIOGRAFÍA COMPLEMENTARIA

- Chen, X., Wills, G. B., Gilbert, L., Bacigalupo, D.: TecRes Report: Using Cloud for Research: a Technical Review. In: Computing, pp. 1--52. UK. (2010).
- Doelitzcher, F., Held, M., Sulistio, A., Reich, C. ViteraaS: Virtual Cluster as a Service. In: 3rd IEEE International Conference on Cloud Computing Technology and Science. Atenas, Grecia (2011).
- Amazon Elastic Compute Cloud (Amazon EC2). <http://aws.amazon.com/es/ec2/>. Febrero 2013.
- Eucalyptus: Open Source software for building AWS-compatible private and hybrid clouds. <http://www.eucalyptus.com>. Febrero 2013.
- Popek, G.J., Goldberg, R.P.: Formal Requirements for Virtualizable Third Generation Architectures. In: Communications in the ACM, Volume 17, Number 7, pp. 412--421. USA. (1974)
- Hacker, T., Mahadik, K. "Flexible Resource Allocation for Reliable Virtual Cluster Computing." In: Supercomputing Proceedings (SC11). Seattle, USA (2011).
- Kondo, D., Javadi, B., Malecot, P., Cappello, F., Anderson, D. P.: "Cost-benefit analysis of Cloud Computing versus desktop grids". In: IPDPS '09 Proceedings. IEEE International Symposium on Parallel and Distributed Processing. Washington, USA (2009).
- Nurmi, D., Wolski, R., Grzegorzczak, C., Obertelli, G., Soman, S., Youseff, L., Zagorodnov, D.: The Eucalyptus Open-Source Cloud-Computing System. In: 9th IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGRID '09), pp. 124-131, IEEE Computer Society, Washington (2009).
- Nussbaum, L., Anhalt, F., Olivier, M., Gelas, J.: Linux-based virtualization for HPC clusters. In: Montreal Linux Symposium (2009), pp. 221—234. Canada. (2009).
- OpenNebula: Open Source Data Center Virtualization. <http://opennebula.org>. Febrero 2013.
- OpenStack Cloud Software: Open source software for building private and public clouds. <http://www.openstack.org>. Febrero 2013.
- Pettoruti, J.E., Rodriguez, I., Chichizola, F., De Giusti, A.: Análisis de la degradación de las comunicaciones en algoritmos de cómputo científico en un Cloud privado. In: Proceedings del XII Workshop de Procesamiento Distribuido y Paralelo (WPDP) - XVIII Congreso Argentino de Ciencias de la Computación (CACIC 2012). Bahía Blanca, Argentina. (2012).
- RackSpace Open Cloud Company. <http://www.rackspace.com>. Febrero 2013.
- Rodriguez, I. P., Pousa, A., Pettoruti, J.E., Chichizola, F., Naiouf, M., De Giusti, L., De Giusti, A.: Estudio del overhead en la migración de algoritmos paralelos de cluster y multicluster a GRID. In: XIII° Congreso Argentino de Ciencias de la Computación (CACIC 2007) Proceedings. Argentina. (2007)
- Rodriguez, I., Pettoruti, J.E., Chichizola, F., De Giusti, A.: Despliegue de un Cloud Privado para entornos de cómputo científico. In: Proceedings del XI Workshop de Procesamiento Distribuido y Paralelo (WPDP) - XVII Congreso Argentino de Ciencias de la Computación (CACIC 2011). La Plata, Argentina. (2011).



**UNIVERSIDAD NACIONAL DE LA PLATA
FACULTAD DE INFORMÁTICA**

- Shafer, J.: "I/O virtualization bottlenecks in cloud computing today". In: Proceedings of the 2nd conference on I/O virtualization (VIOV10). California, USA (2010).
- Vaquero, L. M., Rodero-Merino, L., Caceres, J., Lindner, M.: A Break in the Clouds: Towards a Cloud Definition. In: ACM SIGCOMM Computer Communication Review, Volume 39, Issue 1, pp. 50--55. USA. (2009).
- Vázquez Blanco, B., Huedo, E., Montero, R. S., Llorente, I. M.: "Elastic Management of Cluster-based Services in the Cloud". In: Proceedings pp.19-24, ACM Digital Library. ISBN 978-1-60558-564-2. (2009).
- Velte, A.T., Velte, T.J., Elsenpeter, R.: Cloud Computing: A Practical Approach, McGraw Hill Professional. (2009)
- Xing, Y., Zhan, Y.: "Virtualization and Cloud Computing". In: Proceedings pp.305-312, Springer Link. ISBN 978-3-642-27323-0. (2012).

CONTACTO CON LA CÁTEDRA

- Toda la información de la cátedra, durante la cursada, se refleja en el entorno virtual de enseñanza y aprendizaje WebUNLP (<http://webunlp.unlp.edu.ar>), donde los alumnos tendrán acceso a través de un nombre de usuario y una clave, para trabajar con sus docentes y compañeros de curso.
- Los alumnos podrán consultar temas específicos de su turno de práctica o teoría, mediante la herramienta de mensajería provista en el curso sobre WebUNLP.
- Los docentes de la cátedra atenderán personalmente consultas durante sus clases.