

# HPC en la Nube para el procesamiento de datos: Simulación y Despliegue en Arquitecturas IaaS.

Diego Encinas<sup>1,2</sup>, Brian Galarza<sup>1</sup>, Román Bond<sup>1</sup>, Gonzalo Zaccardi<sup>1</sup>, Nicolás Benquerença Mendes<sup>1</sup>, Rodrigo Lago<sup>1</sup>, Cristian Iglesias<sup>1</sup>, Rodrigo Iván Duarte<sup>1</sup>, Jorge Osio<sup>1</sup>, David Duarte<sup>1</sup>, Martín Morales<sup>1,3</sup>

<sup>1</sup>Proyecto de Investigación SimHPC - Programa TICAPPS - Instituto de Ingeniería y Agronomía - UNAJ

<sup>2</sup>Instituto de Investigación en Informática LIDI (III-LIDI) - Facultad de Informática - UNLP – Centro Asociado CIC

<sup>3</sup>Centro CodApli - Facultad Regional La Plata - UTN

dencinas@unaj.edu.ar, bgalarza@unaj.edu.ar, rbond@unaj.edu.ar, gzaccardi@unaj.edu.ar, nicobenquerenca@yahoo.com.ar, rodrigo498@gmail.com, rodrigo4@hotmail.com, josio@unaj.edu.ar, davito.duarte.22@gmail.com, martin.morales@unaj.edu.ar



## Resumen

El objetivo de esta línea de investigación es el estudio de la performance de las arquitecturas tipo cloud a través del despliegue de IaaS y utilización de IaaS públicos, en particular en el área de cómputo paralelo de altas prestaciones (HPC). Enfocando a la obtención de herramientas que permitan

predecir la eficiencia del sistema ante posibles escenarios.

## Contexto

Se presenta una línea de Investigación que es parte del Proyecto de Investigación "Simulación, Computación de Altas Prestaciones (HPC) y optimización de aplicaciones sociales – SimHPC" de la Universidad Nacional Arturo Jauretche (UNAJ), acreditado por resolución interna 183/21. Además, el proyecto colabora con el Programa "Tecnologías de la información y la comunicación (TIC) en aplicaciones de interés social" – TICAPPS de la UNAJ.

## Introducción

Cloud Computing es un paradigma que está en constante crecimiento durante estos últimos años, cada vez más compañías y grupos de investigación trabajan en conjunto con el fin de explotar las oportunidades ofrecidas por el mismo. Dicho paradigma ofrece muchas ventajas, tales como el bajo costo de implementación, ya que no se necesitan computadoras de última tecnología debido a que éstas trabajan conjuntamente (Clustering) con la posibilidad de escalar horizontalmente de manera sencilla. Además, hay software Open Source disponible para los nodos en el clúster como las infraestructuras Eucalyptus, OpenNebula, CloudStack u OpenStack integradas con GNU/Linux y compatibles, por ejemplo, con Amazon WebServices.

## Líneas de investigación y desarrollo

### Temas de Estudio e Investigación

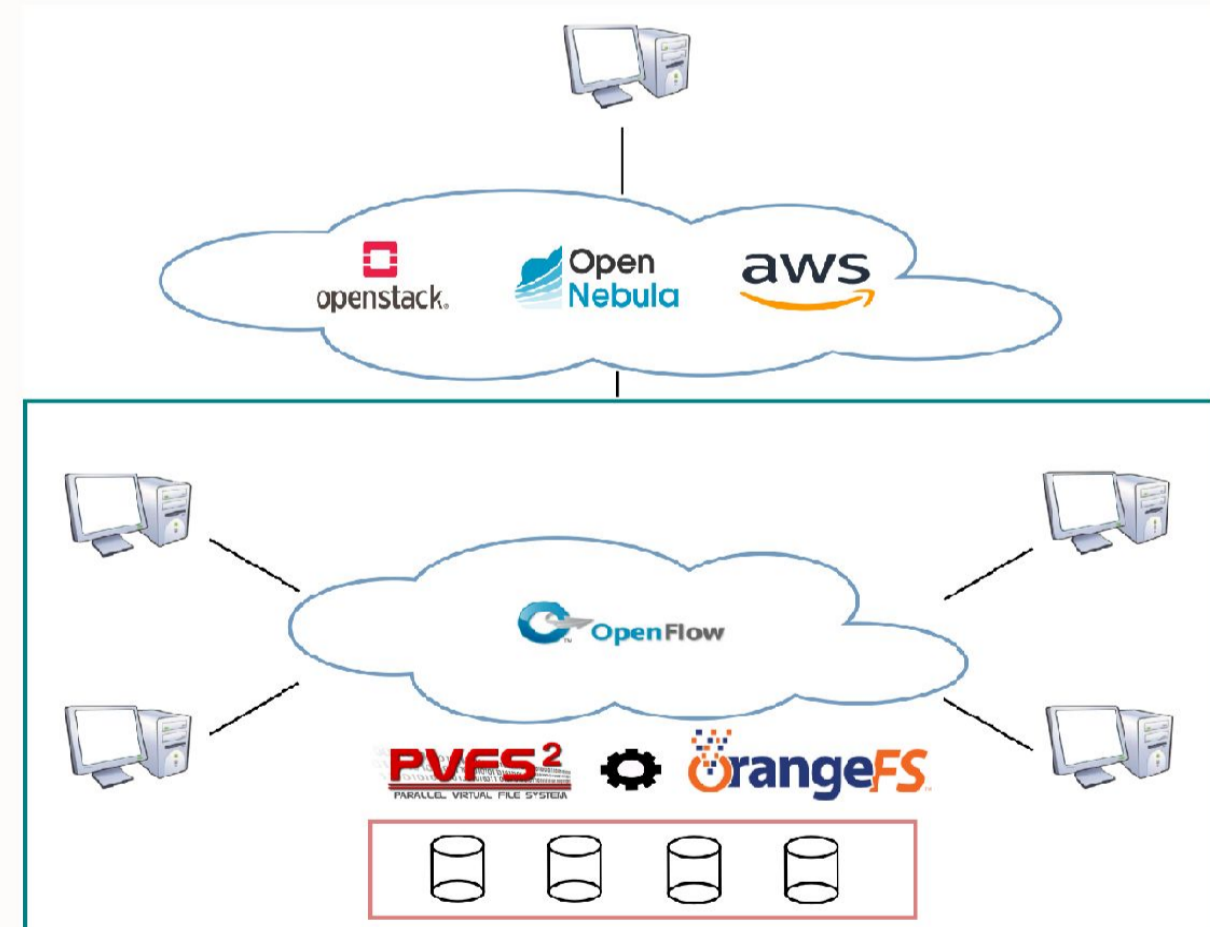
- Arquitecturas multiprocesador para procesamiento paralelo: multiprocesador de memoria compartida, multiprocesador on-chip de memoria distribuida. Multicore, Clusters, Clusters de multicore. Grid. Cloud.
- Plataformas de software para implementar y administrar Clouds públicos, privados e híbridos.
- Sistemas de Archivos Paralelos.
- Procesamientos de Grandes Volúmenes de datos en entornos cloud con herramientas del ecosistema Apache Hadoop y afines.

## Resultados y objetivos

### Investigación experimental

- Implementación de un IaaS encargado de realizar operaciones en procesamiento paralelo aumentando la eficiencia y reduciendo los costes generados.
- Implementación de OpenStack Dashboard y de un sistema desarrollado para poder controlar/administrar de manera visual (web) y más básica cada uno de los servicios.
- Implementación de OpenNebula en un sistema con las mismas características que el implementado por OpenStack con el fin de poder realizar pruebas en entornos similares.
- Utilización de Fuel para administrar OpenStack como sistema de administración de nube (Cloud Computing) a partir de la infraestructura de 2 nodos compute y el controller.

- Lograr escalabilidad agregando nuevos nodos compute a la infraestructura obteniendo un mayor performance en el sistema.
- Análisis del rendimiento de un Cloud privado en la ejecución de instancias personalizadas.
- Ejecutar sobre el modelo de nodos implementado en OpenStack la distribución GNU/Linux Hetnux, desarrollada en la UNAJ.
- Ejecutar diferentes benchmarks en la infraestructura desplegada sobre OpenStack y OpenNebula para así comparar los resultados obtenidos de ambas infraestructuras y poder realizar un análisis del rendimiento en cada caso.
- Análisis y configuración de clusters virtuales.
- Análisis y configuración de herramientas no invasivas para la obtención de métricas en las distintas capas de software de los sistemas de archivos paralelos.
- Utilización de otros sistemas de archivos paralelos como Lustre y Beegfs para obtener métricas en Metadataservidores.



## Formación de Recursos Humanos

Durante 2025 se han realizado 5 publicaciones nacionales: una en el Congreso Argentino de Ciencias de la Computación (CACIC) y cuatro publicaciones en el Congreso Nacional de Ingeniería Informática / Sistemas de Información (CoNalISI). En esta línea de I/D existe cooperación a nivel nacional. Hay 3 investigadores realizando carreras de postgrado y alumnos avanzados de grado colaborando en las tareas.

Además, alumnos de las materias Sistemas Operativos 1, Redes de Computadoras 2, Programación en Tiempo Real y Organización y Arquitecturas de Computadoras colaboran con trabajos y tareas de investigación. Por otro lado, algunos integrantes participan en el dictado de la Maestría y Diplomatura en Ciencia de Datos de la UNAJ.