

Introducción al Cómputo de Alto Rendimiento (HPC) para Economía y Ciencias Sociales

Ignacio Sarmiento-Barbieri & Juan Ernesto Sanchez

Universidad de los Andes

¿Qué es el HPC y cómo funciona?

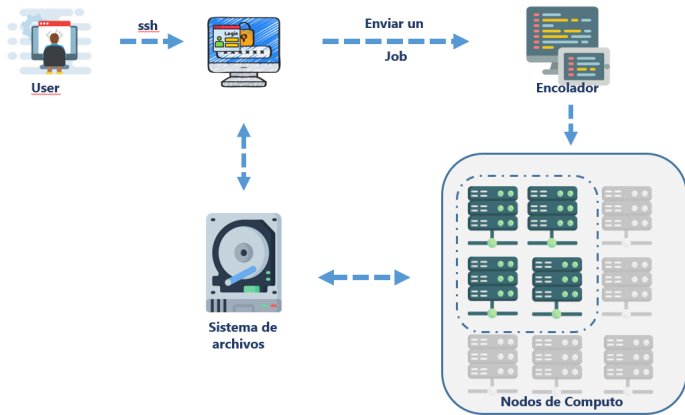


Figure 1: HPC

Objetivo del curso

Propósito y audiencia

- ▶ Curso modular enfocado en el uso del sistema HPC enfocado a los problemas de nuestros usuarios.
- ▶ Dirigido a maestría, doctorado, asistentes de investigación y profesores (¿educación continua?).
- ▶ Integrar el clúster Hypatia en flujos de trabajo empíricos.

Diseño del curso

Estructura modular

- ▶ Cubre desde la conexión y Linux hasta SLURM y entornos de software.
- ▶ Incluye ejemplos de estimación típicos de nuestros usuarios.
- ▶ Módulos con ejercicios representativos y un proyecto final de autonomía en HPC.

Metodología de enseñanza

Materiales y evaluación formativa

- ▶ Es un curso asincrónico
- ▶ Videos y guías paso a paso.
- ▶ Preguntas de reflexión o autoevaluación por módulo.
- ▶ Preparación para la evaluación final, aprendizaje al propio ritmo.

Módulo 1, Introducción al HPC

Objetivo y contexto

- ▶ Visión general del HPC y del clúster Hypatia.
- ▶ Desmitificar el uso, beneficios y bases conceptuales para los módulos posteriores.
- ▶ ¿Qué tareas conviene ejecutar en el clúster y cuáles no?

Módulo 2, Conexión al HPC

SSH para Windows y Mac

- ▶ Conectarse al sistema mediante SSH con instrucciones por sistema operativo.
- ▶ Transición amigable para usuarios de entornos gráficos.
- ▶ Lograr conexiones estables, seguras y reproducibles.

Módulo 3, Navegación Linux

Estructura de directorios y comandos básicos

- ▶ Fluidez en terminal para moverse en el clúster.
- ▶ Lógica jerárquica, rutas absolutas y relativas.
- ▶ Operaciones básicas con archivos y carpetas.

Módulo 4, Trabajo remoto y transferencia

Integración local y remota

- ▶ Transferir archivos y editar scripts de forma remota.
- ▶ Opciones desde básicas a avanzadas según experiencia.
- ▶ Integración con editores y montaje de directorios.

Módulo 5, Uso de SLURM

Ejecución de trabajos

- ▶ SLURM como planificador para gestionar recursos.
- ▶ Sesiones interactivas para exploración y prueba.
- ▶ Trabajos batch para cargas largas y reproducibles.
- ▶ Estructura básica de scripts SLURM y variables de entorno.

Módulo 6, Software y entornos

Módulos, venvs y contenedores

- ▶ Sistema de módulos para cargar versiones específicas de software.
- ▶ Evitar conflictos de dependencias y automatizar configuraciones.
- ▶ Entornos virtuales en R y Python, e introducción a contenedores.

Módulo 7, Ejemplos prácticos

Flujo integral de trabajo en HPC

- ▶ Estimar recursos, configurar entorno, ejecutar y analizar resultados.
- ▶ Enfoque en decisiones reales de implementación.
- ▶ Criterios para paralelizar y dimensionar memoria y CPU.

Módulo 7, Python

CPU versus GPU

- ▶ **Random Forest** con `scikit-learn`, uso de `n_jobs`.
- ▶ **Red neuronal** con Keras o PyTorch, comparación CPU y GPU.
- ▶ Identificar cuándo el uso de GPU justifica el costo.

Módulo 7, R

Econometría a escala

- ▶ **Efectos fijos:** `lm(y ~ i.firma + i.year)`, `reghdfe` [Stata], `fixest::feols` [R].
- ▶ Discutir tiempos, RAM, algoritmos [FWL, QR] y eficiencia.
- ▶ **DiD** con `did::att_gt`, escalamiento y panel eficiente.
- ▶ **Bootstrap y BoB** para paralelizar remuestreo y reducir costos.

Módulo 7, Más allá de R y Python

Bajo nivel y optimización

- ▶ Fortran y C++ para maximizar velocidad y eficiencia.
- ▶ Detectar cuellos de botella en código de alto nivel.
- ▶ Cuándo usar bibliotecas optimizadas y cuándo reescribir en C++ o Fortran.

Módulo 8, Complejidad y recursos

Estimación razonada

- ▶ Complejidad algorítmica, FLOPS, memoria y tiempo de ejecución.
- ▶ Medir objetos en memoria y tiempos de bloques de código.
- ▶ Extrapolar para dimensionar correctamente trabajos en SLURM.

Módulo 9, Proyecto práctico

Evaluación integral

- ▶ Implementar un flujo completo de trabajo en el clúster.
- ▶ Uso adecuado de recursos, gestión de archivos y módulos.
- ▶ Monitoreo correcto de la ejecución.

Módulo 9, Estructura del proyecto

Pasos del flujo

- ▶ Descargar datos desde Bloque Neón al equipo local.
- ▶ Subir archivos al clúster [SCP, SFTP, etc.].
- ▶ Configurar entorno [módulos, scripts].
- ▶ Ejecutar en SLURM en modo batch.
- ▶ Descargar resultados y visualizarlos localmente.

Módulo 9, Ejemplos de proyectos

Aplicaciones sugeridas

- ▶ **Random Forest** para clasificación o regresión en datos estructurados.
- ▶ **Efectos fijos** en R o Stata con bootstrap de errores estándar aprovechando SLURM.

Módulo 9, Entregables

Validación y documentación

- ▶ Quiz breve de validación en Bloque Neón.
- ▶ Script de envío SLURM .sh claro y comentado.
- ▶ Salidas del código para verificación.

Módulo 10, Buenas prácticas

Recomendaciones avanzadas

- ▶ Optimizar trabajo, mantener orden y aprovechar entorno compartido.
- ▶ Scripts base y recursos del curso para consulta posterior.
- ▶ Enfocado en quienes desean profundizar.