

Cluster Argentino de Bajo Costo para Molecular Dynamics y Machine Learning

Resumen ejecutivo: Con una inversión de ochenta mil dólares estadounidenses es posible armar un cluster de 36 GPUs de última generación para dinámica molecular y aprendizaje automático. Más de cien investigadoras/es de primera línea del sistema de CyT del país podrán usarlo y mejorar sus herramientas de investigación. La relación MFLOPS/USD es cinco veces mejor que la mejor compra de equipamiento de cómputo que se haya hecho en Ciencia y Técnica en los últimos años.

Desde el Centro de Computación de Alto Desempeño (CCAD) de la Universidad Nacional de Córdoba (UNC) estamos impulsando una inversión en tarjetas gráficas para realizar cómputo de dinámica molecular y aprendizaje automático.

El CCAD realizó en 2017 una **experiencia exitosa en un reposicionamiento de hardware**, transformando dos nodos del viejo cluster Mendieta (<https://ccad.unc.edu.ar/equipamiento/cluster-mendieta/>) en computadoras con tarjetas gráficas (GPUs) para cálculos. Estos dos nodos se llaman Nabucodonosor (<https://ccad.unc.edu.ar/equipamiento/computadora-nabucodonosor/>) y desde que están en funcionamiento comprobamos que tienen un excelente desempeño en *workloads* de MD (*molecular dynamics*) y ML (*machine learning*). A tal punto resultan exitosos, que muchas investigadoras e investigadores de todo el país han requerido cuenta para poder correr en estos dos *workloads*, *a priori*, disímiles.

Hemos realizado un **relevamiento de personas y grupos de investigación interesados en contar con lo que sería el Primer Clúster Nacional de GPUs**. Tenemos casi **un centenar de personas de ciencia y técnica de todo el país** (Universidades e Institutos de CONICET) con necesidad de correr en un cluster de estas características para mejorar la calidad de sus investigaciones. Más corridas, modelos más detallados, mayor rango de tiempo de simulación.

Los costos para su adquisición son bajos, ya que gran parte de la inversión está hecha por el CCAD-UNC y son los nodos de Mendieta. Solo resta comprar las GPUs y pagar el servicio de re-adequación de los servidores. Teniendo en cuenta que cada GPU sale alrededor de 2100 USD, que cada nodo incorpora tres (3) GPUs y el costo de re-adequación de un nodo es de 400 USD, **reconvertir los 12 nodos de Mendieta implican una inversión de 80000 USD**. Con esto lograríamos un rendimiento pico teórico de 360 TFLOPS de precisión simple *-float32-* que es la precisión más utilizada en los *workloads* de MD y ML. Esto implica una relación de 4500 MFLOPS/USD, que es cinco veces mejor que una excelente compra que hizo la UNC en septiembre de 2020, logrando 856 MFLOPS/USD (*float32*, pico) para su nuevo cluster Serafín.

¿Por qué una computadora para Machine Learning sirve para Molecular Dynamics y viceversa?

En *High Performance Computing* (HPC) se habla de **workloads similares** cuando en dos computaciones se utilizan con igual balance ancho de banda de memoria, unidades de procesamiento de punto flotante, unidades de procesamiento de punto entero y paralelismo.

Por un lado tenemos *Machine Learning* (ML) que desde siempre se ha promocionado como el nicho donde las Graphics Processing Units (GPUs) resultan muy convenientes respecto a la relación de cálculos/tiempo (velocidad) y cálculos/inversión (eficiencia económica). Esto es así porque las GPUs presentan una forma de paralelismo, ancho de banda a memoria y tipos de datos que es muy convenientes para este *workload* que realiza muchos cálculos de punto flotante de simple y media precisión (*float32*, *float16*, *bfloat16*, *tfloat16*, *float24*).

Por otra parte en *Molecular Dynamics* (MD), se empezó a utilizar cada vez más una técnica que se denomina *mixed precision* [1], para aprovechar el desbalance entre unidades de cómputo *float64* y *float32* que hay en las GPUs *gamers* (1:8, 1:32) respecto al diseño tradicional en CPU (1:2), además de aprovechar mejor la memoria en tamaño y ancho de banda, ya que para representar la posición de un átomo alcanza con *float32*. Lentamente todos los paquetes de MD (AMBER, GROMACS, LAMMPS, etc.) empezaron a incorporar esta técnica y actualmente, la gran mayoría de ellos hace casi todos sus cálculos en *float32* [2]. Este alejamiento de la precisión doble es una tendencia general [3].

Una GPU gamer es una decisión de diseño que implica utilizar gran parte de los ~400nm² de silicio para transistores de ~10nm en unidades de cómputo para *float32*, ya que ese es el tipo de datos que los videojuegos utilizan para lograr el realismo 3D que la industria requiere.

Los *workloads* de ML y MD son, a los ojos de las y los usuarios, dos cargas de trabajo en la cual la GPU brilla, tanto por su velocidad en TFLOPS, como por su conveniencia económica, TFLOPS/USD.

Dr. Diego Masone, Dr. Nicolás Wolovick
para el

Centro de Computación de Alto Desempeño (CCAD)
Universidad Nacional de Córdoba (UNC)
Octubre, 2020

Referencias

- [1] Scott Le Grand, Andreas W. Götz, Ross C. Walker, "[SPFP: Speed without compromise—A mixed precision model for GPU accelerated molecular dynamics simulations](#)", Computer Physics Communications, 184(2), 2013.
- [2] Carsten Kutzner, Szilárd Páll, Martin Fechner, Ansgar Esztermann, Bert L. de Groot, Helmut Grubmüller, "[More bang for your buck: Improved use of GPU nodes for GROMACS 2018](#)", Journal of Computational Chemistry, 40(27), 2019.
- [3] J. Domke et al., "[Double-Precision FPUs in High-Performance Computing: An Embarrassment of Riches?](#)", 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS), Rio de Janeiro, Brazil, 2019, pp. 78-88, doi: 10.1109/IPDPS.2019.00019.

Firmantes

Aprendizaje Automático (ML)

Científico/a	Institución	Email
MAGUITMAN, Ana	ICIC	agm@cs.uns.edu.ar
PONZONI, Ignacio	ICIC	ip@cs.uns.edu.ar
SOTO, Axel	ICIC	saj@cs.uns.edu.ar
DOMINGUEZ, Mariano Javier	IATE	mardom@oac.unc.edu.ar
FERRANTE, Enzo	sinc(i)	eferrante@sinc.unl.edu.ar
NIETO, Nicolás	sinc(i)	nnieto@sinc.unl.edu.ar
MILONE, Diego	sinc(i)	dmilone@sinc.unl.edu.ar
ECHEVESTE, Rodrigo	sinc(i)	recheveste@sinc.unl.edu.ar
YONES, Cristian	sinc(i)	cyones@sinc.unl.edu.ar
BUGNON, Leandro	sinc(i)	lbugnon@sinc.unl.edu.ar
GERARD, Matias	sinc(i)	mgerard@sinc.unl.edu.ar
STEGMAYER, Georgina	sinc(i)	gstegmayer@sinc.unl.edu.ar
TORIANO, Roxana Mabel	IFIBIO Houssay	rtoriano@fmed.uba.ar
DELRIEUX, Claudio	LCI	cad@uns.edu.ar
THOMSEN, Felix	LCI	felix.thomsen@uns.edu.ar
REVOLLO, Natalia	LCI	revollonatalia@gmail.com
REVOLLO, Noelia	LCI	giselarevollo@gmail.com
CIPOLLETTI, Marina	LCI	mpcipolletti@gmail.com

LOMBARDI, Leandro Ezequiel	IC	llombard@dm.uba.ar
ACIÓN, Laura	IC	lacion@gmail.com
FERRER, Luciana	ICC	lferrer@dc.uba.ar
TAMARIT, Francisco	IFEG	francisco.tamarit@unc.edu.ar
RUDERMAN, Andrés	IFEG	ruderman@famaf.unc.edu.ar
PORTA, Juan	IFEG	jporta.09@gmail.com
SANCHEZ, Jorge	CIEM	jsanchez@famaf.unc.edu.ar
BENOTTI, Luciana	FaMAF	luciana.benotti@unc.edu.ar
LUQUE, Franco	FaMAF	francolq@unc.edu.ar
TERUEL, Milagro	FaMAF	milagro.teruel@gmail.com
DOMÍNGUEZ, Martín	FaMAF	mdoming@famaf.unc.edu.ar
ALONSO ALEMANY, Laura	FaMAF	lauraalonsoalemány@gmail.com
CARDELLINO, Cristian	FaMAF	crscardellino@gmail.com
CELAYES, Pablo	FaMAF	pablocelayes@gmail.com
SERRANO, Facundo	FaMAF	facu64_17@hotmail.com
MERILES, Emanuel	FaMAF	emeriles@gmail.com
FIURI, Ariel	FaMAF	ariel.fiuri@gmail.com
BARSCE, Juan Cruz	UTN-FRVM	jbarsce@gmail.com
PALOMBARINI, Jorge	UTN-FRVM	palombarini@gmail.com
BECCARIA, Ezequiel	UTN-FRVM	ezequielbeccaria@gmail.com

Dinámica Molecular (MD)

Científica/o	Institución	Email
MARISCAL Marcelo	INFIQC	marrmariscal@fcq.unc.edu.ar
PATRITO, Martín	INFIQC	mpatrilo@gmail.com
DE LA ROSA ABAD, Andrés	INFIQC	juanderoab@hotmail.com
PAREDES, Patricia	INFIQC	patricia@fcq.unc.edu.ar
FARIGLIANO, Lucas Martin	INFIQC	lfarigliano@gmail.com
SORIA, Federico Ariel	INFIQC	fedesoria1@gmail.com

OVIEDO, Belén	INFIQC	beloviedo@gmail.com
PAZ, Alexis	INFIQC	alexis.paz@gmail.com
JIMENEZ, Juan	INFIQC	juanjimenezg92@gmail.com
OLMOS ASAR, Jimena Anahi	INFIQC	jimena.olmos.asar@unc.edu.ar
PUIATTI, Marcelo	INFIQC	marcelo.puiatti@gmail.com
REINAUDI, Luis	INFIQC	luis.reinaudi@unc.edu.ar
ESTRIN, Dario Ariel	INQUIMAE	dario@qi.fcen.uba.ar
SCHERLIS PEREL, Damián Ariel	INQUIMAE	damian@qi.fcen.uba.ar
CAPECE, Luciana	INQUIMAE	lula@qi.fcen.uba.ar
ARRAR, Mehrnoosh	INQUIMAE	mehrnoosh.arrar@gmail.com
LARIA, Daniel Hector	INQUIMAE	dhlaria@cnea.gov.ar
FACTOROVICH, Matias Hector	INQUIMAE	mfactorovich@qi.fcen.uba.ar
RODRIGUEZ, Javier	UE-INN	javier@speedy.cnea.gov.ar
ELOLA, Maria Dolores	UE-INN	DoloresElola@gmail.com
VILA, Jorge Alberto	IMASL	jorgevila84@gmail.com
PORASSO, Rodolfo Daniel	IMASL	rporasso@unsl.edu.ar
PEREYRA, Rodolfo Guillermo	IFEG	pereyra@famaf.unc.edu.ar
GIMENEZ, Maria Cecilia	IFEG	cgimenez@famaf.unc.edu.ar
SERRA, Pablo	IFEG	pablo.serra@unc.edu.ar
TURJANSKI, Adrian	IQUIBICEN	adrian@qb.fcen.uba.ar
MARTI, Marcelo Adrian	IQUIBICEN	marti.marcelo@gmail.com
FREIBERGER, Maria Ines	IQUIBICEN	mariaines.freiberger@gmail.com
DRUSIN, Salvador	IQUIBICEN	sdrusin@gmail.com
FERNANDEZ ALBERTI, Sebastian	UNQ	sfalberti@gmail.com
PALMA, Juliana Isabel	UNQ	juliana@unq.edu.ar
PIERDOMINICI SOTTILE, Gustavo	UNQ	gsottile@unq.edu.ar
BALATTI, Galo Ezequiel	UNQ	gbalatti@df.uba.ar
ENRIZ, Ricardo Daniel	IMIBIO-SL	denriz@unsl.edu.ar

VEGA HISSI, Esteban Gabriel	IMIBIO-SL	egvega@unsl.edu.ar
ZARYCZ, Maria Natalia Cristina	IMIBIO-SL	nzarycz@conicet.gov.ar
ANDUJAR, Sebastián	IMIBIO-SL	saanduja@unsl.edu.ar
DEL POPOLO, Mario	ICB	mdelpopolo@gmail.com
RUESTES, Carlos	ICB	cjruestes@hotmail.com
SOSA, Micaela	ICB	smicaelajanet@gmail.com
HOYUELOS, Miguel	IFIMAR	hoyuelos@mdp.edu.ar
DI MURO, Matías	IFIMAR	mdimuro@mdp.edu.ar
DALOSTO, Sergio Daniel	IFIS - LITORAL	sergio.dalosto@santafe-conicet.gov.ar
TINTE, Silvia	IFIS - LITORAL	silvia.tinte@santafe-conicet.gov.ar
BRINGA, Eduardo	UM	ebringa@yahoo.com
APARICIO, Emiliano Jorge	UM	emilianojaparicio@gmail.com
APARICIO, Romina	UM	romi.a.arias@gmail.com
DOS SANTOS, Gonzalo	UM	gonzalodossantos@gmail.com
DELUIGI, Orlando Raul	UM	o.r.deluigi@gmail.com
ANGELINA, Emilio	IQUIBA-NEA	emilioluisangelina@hotmail.com
CAVASOTTO, Claudio Norberto	IIMT	ccavasotto@austral.edu.ar
MASONE, Diego	IHEM	dmasone@mendoza-conicet.gob.ar
MILLÁN, Emmanuel	ITIC	emmanueln@gmail.com
FERNANDEZ, María Laura	INFIP	mlaura@df.uba.ar
FERRERO, Ezequiel	INN-CAB	ferrero@famaf.unc.edu.ar
MARTINI, Florencia	IQUIMEFA	flormartini1@gmail.com
ÁVILA, César	INSIBIO	clavila@gmail.com
CLEMENTE, Camila	UNVM	camilamaraclemente@gmail.com
BOECHI, Leonardo	IC	lboechi@ic.fcen.uba.ar
SANCHEZ, Veronica Muriel	CSC	veronica.sanchez@csc.conicet.gov.ar
PICKHOLZ, Monica Andrea	IFIBA	mpickholz@gmail.com
PASTORINO, Claudio	UE-INN	pastor@cnea.gov.ar
GALIGNIANA, Mario Daniel	IBYME	mgaligniana@conicet.gov.ar

RODRIGUES, Daniel Enrique	INTEC	danielr@fcb.unl.edu.ar
CARLEVARO, Carlos Manuel	IFLYSIB	manuel@iflysib.unlp.edu.ar
GIL REBAZA, Arles Victor	IFLP	arlesv@fisica.unlp.edu.ar
VERA, Domingo Mariano Adolfo	INBIOTEC-QUIA MM	dmavera@yahoo.com
TEN HAVE, Arjen	IIB	tenhave.arjen@gmail.com
MONTICH, Guillermo Gabriel	CIQUIBIC	gmontich@fcq.unc.edu.ar
SIANO, Alvaro Sebastian	UNL	asiano@fcb.unl.edu.ar
MIGUEL, Virginia	IIBYT	virgimiguel@yahoo.com.ar
QUEVEDO, Alfredo	UNITEFA	aquevedo@unc.edu.ar
ADLER, Natalia	CIBION	natu.adler@gmail.com
MARINO BUSLJE, Cristina	IIBBA	cmb@leloir.org.ar
BALBUENA, Cristian	INTEMA	cbalbuena@fi.mdp.edu.ar