Julio López

julio.lopez@cmu.edu
http://www.cs.cmu.edu/~jclopez
5000 Forbes Ave, Pittsburgh PA 15213

Professional Interests

Systems and applications for *data intensive computing at large scale*, including computational databases, parallel and distributed systems, scalable I/O and indexing techniques for large multi-dimensional spatial datasets, data compression and visualization. In particular, I am interested in programming models, abstractions and supporting systems for these types of computation.

Education

Ph.D. Electrical and Computer Engineering,
Carnegie Mellon University, Pittsburgh PA, USA. Aug 2007.
Thesis: "*Methods for Querying Compressed Wavefields*". Techniques to compress, index and query large wavefield datasets in their compressed representation, turning an I/O intensive problem into a massively parallel computational workload.

M.Sc. Electrical and Computer Engineering,
Carnegie Mellon University, Pittsburgh PA, USA. May 2000.
Thesis: "*Dv — A framework for remote visualization*". Framework to create visual representations, such as images or 3D models, of datasets stored at remote sites.

B.S. Computer Science (Ingeniero de Sistemas).
Universidad EAFIT, Medellín, Colombia. June 1996.
Graduated at the top of the class. Highest GPA for the class every term. Graduation Project: "*A Protocol for Communications in Distributed Virtual Environments*". Communication layer of a larger project intending to implement a prototype Distributed Virtual Reality system.

Professional Experience

Carnegie Mellon University. Pittsburgh, PA, USA. Oct 2007 – present.
Systems Scientist, Parallel Data Laboratory. Research in data-intensive, cloud computing and eScience. System architect and technical leader in the deployment of the CMU PDL data-intensive clusters (Open Cloud and Open Cirrus).

Hewlett-Packard Laboratories. Palo Alto CA, USA. June 2000 – September 2000.
Intern Researcher: Designed and implemented a mechanism to aggregate the bandwidth of the long-range links of multiple personal devices.

Massachusetts Institute of Technology (MIT)/Center for Educational Computing Initiatives.
Boston MA, USA. August 1997 – July 1998.
Visiting Researcher: Developed a library to annotate, classify, and store multimedia objects. Designed and implemented a caching architecture for low bandwidth networks.

Universidad EAFIT, Proyecto Conexiones. Medellín, Colombia. January 1997 – July 1997.
Software Engineer: Ported to Java the user interface of the "*Pachamama*" educational software.

Universidad EAFIT, IT Department. Medellín, Colombia. June 1995 – January 1997
System and Network Administrator: Administered several Solaris / HP-UX / Linux / MS Windows machines and Cisco routers. Configured, deployed and managed numerous network services software for these environments.

Corporación Nacional de Ahorro y Vivienda (CONAVI). Medellín, Colombia. January 1995–June 1995.
EDP Auditor: Corporate network performance and security evaluation.

Publications

Kai Ren, Julio López and Garth Gibson. "*Otus: Resource Attribution and Metrics Correlation in Data-Intensive Clusters*". MapReduce: The Second International Workshop on MapReduce and its Applications. San Jose, CA, June 2011.

Julio López, Colin Degraf, Tiziana DiMatteo, Bin Fu, Eugene Fink and Garth Gibson, "*Recipes for Baking Black Forest Databases: Building and Querying Black Hole Merger Trees from Cosmological Simulations*". Submitted to SSDBM 2011.

Ricardo Taborda, Julio López, Haydar Karaoglu, John Urbanic, Jacobo Bielak. "*Speeding Up Finite Element Wave Propagation for Large-Scale Earthquake Simulations*". Technical Report CMU-PDL-10-109, Parallel Data Laboratory, Carnegie Mellon University, October 2010.

Bin Fu, Kai Ren, Julio López, Eugene Fink and Garth Gibson. "*DiscFinder: A data-intensive scalable cluster finder for astrophysics (extended version)*". Technical Report CMU-PDL-10-104, Parallel Data Laboratory, Carnegie Mellon University, October 2010.

Bin Fu, Kai Ren, Julio López, Eugene Fink, and Garth Gibson. "*DiscFinder: A data-intensive scalable cluster finder for astrophysics*". In Proceedings of the ACM International Symposium on High Performance Distributed Computing (HPDC), Chicago, IL. June, 2010.

Julio López, Leonardo Ramirez-Guzman, Jacobo Bielak, and David O'Hallaron. "*BEMC: A Searchable, Compressed Representation for Large Seismic Wavefields*". In proceedings of the 22nd International Conference Scientific and Statistical Database Management (SSDBM 2010), Heidelberg, Germany, June 2010.

Swapnil Patil, Garth A. Gibson, Gregory R. Ganger, Julio López, Milo Polte, Wittawat Tantisiroj, and Lin Xiao, "*In Search of an API for Scalable File Systems: Under the Table or Above It?*" In Workshop on Hot Topics in Cloud Computing (HotCloud'09), San Diego CA, June 2009.

Michael A. Kozuch, Michael P. Ryan, Richard Gass, Steven W. Schlosser, David O'Hallaron, James Cipar, Elie Krevat, Michael Stroucken, Julio López, Gregory R. Ganger. "*Tashi: Location-aware Cluster Management*". In proceeding of the Workshop on Automated Control for Datacenters and Clouds (ACDC'09), Barcelona, Spain, June 2009.

Steven Schlosser, Michael Ryan, Ricardo Taborda, Julio López, David O'Hallaron, and Jacobo Bielak, "*Materialized community ground models for large-scale earthquake simulation*", In Proceedings of the Supercomputing Conference (SC'08), Austin, TX, Nov 2008.

Grant Mackey, Saba Sehrish, John Bent, Julio López, Salman Habib, Jun Wang. "*Introducing Map-Reduce to High End Computing*". In Proceedings of the 3rd Petascale Data Storage Workshop, Austin, TX, November 2008.

Julio López. Methods for Querying Compressed Wavefields. Ph.D. Thesis. Electrical and Computer Engineering Department, Carnegie Mellon University. Pittsburgh, PA. May 2007.

Michael Mesnier, Matthew Wachs, Julio López, Raja Sambasivan, James Hendricks and Gregory Ganger. "*‖TRACE – Parallel Trace Replay with Approximate Causal Events*". In Proceedings of 5th Conference on File And Storage Technologies (FAST'07), February 2007.

Leonardo Ramirez-Guzman, Ricardo Taborda, Jacobo Bielak, Omar Ghattas, Eui Joong Kim, Julio López, David O'Hallaron, and Tiankai Tu. "*Scaling up TeraShake: A 1-Hz Case Study*". In EOS Trans. American Geophysical Union (AGU). Volume 87(52), San Francisco, CA, December 2006

Hongfeng Yu, Tiankai Tu, Jacobo Bielak, Omar Ghattas, Julio López, Kwan-Liu Ma, David O.Hallaron,

Leonardo Ramirez-Guzman, Nathan Stone, Ricardo Taborda-Rios, and John Urbanic. "*Remote Runtime Steering of Integrated Terascale Simulation and Visualization*". In Proceedings of the Supercomputing Conference (SC06), Tampa, Fl, November 2006. Winner of the HPC Analytics Challenge at SC06.

Ricardo Taborda, Leonardo Ramirez-Guzman, Tiankai Tu, Julio López, Jacobo Bielak and David O'Hallaron. "*Terashake simulations using Hercules: analysis and comparison.*", In Proceedings of the SCEC Annual Meeting, Palm Springs, CA , September 2006.

Stratos Papadomanolakis, Anastassia Ailamaki, Julio López, Tiankai Tu, David R. O'Hallaron, Gerd Heber. "*Efficient query processing on unstructured tetrahedral meshes*". In Proceedings of the ACM International Conference on Management Of Data (SIGMOD'06), June 2006. pp. 551–562.

Eno Thereska, Brandon Salmon, John Strunk, Matthew Wachs, Michael Abd-El-Malek, Julio López, Gregory R. Ganger. "*Stardust: Tracking activity in a distributed storage system*". In Proceedings of the Joint International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS'06), June 26th–30th 2006.

Minster, J., Olsen, K B., Moore, R., Day, S., Maechling, P., Jordan, T., Faerman, M., Cui, Y., Ely, G., Hu, Y., Shkoller, B., Marcinkovich, C., Bielak, J., Okaya, D., Archuleta, R., Wilkins-Diehr, N., Cutchin, S., Chourasia, A.,Kremenek, G., Jagatheesan, A., Brieger, L., Majundar, A., Chukkapalli, G., Xin, Q., Banister, B., Thorp, D., Kovatch, P., Diegel, L., Sherwin, Thiebaux, M. and López J., "*The SCEC TeraShake Earthquake Simulation*", In EOS Trans. American Geophysical Union (AGU), Volume 47, 2004.

Julio López, David O'Hallaron and Tiankai Tu. "*Big Wins with a Small Application-Aware Cache*". In Proceedings of the Supercomputing Conference (SC'04), November 2004.

Tiankai Tu and David R. O'Hallaron and Julio López. "*Etree: a database-oriented method for generating large octree meshes*". In Journal of Engineering with Computers, Volume 20, Number 2, 2004, pages 117–128.

Volkan Akcelik, Jacobo Bielak, George Biros, Ioannis Epanomeritakis, Antonio Fernandez, Omar Ghattas, Eui Joong Kim, Julio López, David O'Hallaron, Tiankai Tu and John Urbanic. "*High-Resolution Forward and Inverse Earthquake Modeling on Terasacale Computers*". In Proceedings of the Supercomputing Conference (SC'03), 2003.

Tiankai Tu, David R. O'Hallaron, and Julio López. "*The Etree Library: A System for Manipulating Large Octrees on Disk*". Technical Report CMU-CS-03-174, School of Computer Science, Carnegie Mellon University, July, 2003.

Tiankai Tu, David O'Hallaron and Julio López, "*Etree — a database-oriented method for generating large octree meshes*". In Proceedings of the 11th International Meshing Roundtable (IMR-11), September 2002, pp. 127–138.

Julio López and David O'Hallaron "*Evaluation of a resource selection mechanism for complex network services*". In Proceedings of the 10th symposium on High Performance Distributed Computing (HPDC-10), August 2001.

Puneet Sharma, Sung-Ju Lee, Jack Brassil and Julio López. "*Dynamic Link Striping: Breaking the WAN Bandwidth Barrier in Piconets*". In Proceedings of the 11th IEEE Workshop on Local and Metropolitan Area Networks, March 2001.

Julio López and David O'Hallaron. "*Support for interactive heavyweight services*". Technical Report CMU-CS-01-104, School of Computer Science, Carnegie Mellon University, February, 2001.

Julio López and David O'Hallaron. "*Run-time support for adaptive heavyweight services*". In Proceedings of the 5th Workshop on Languages, Compilers and Run-time systems (LCR 2000), May 2000, Vol 1915 of Springer Verlag Lecture Notes in Computer Science, pp. 221–234.

Julio López and Edward Hogan. "*Jiffy: A Lightweight Jini File System*". In Proceedings of 2nd CMU Student Symposium on Computer Systems (SOCS-2), Oct 1999.

Julio López, Martin Aeschlimann, Peter Dinda, Bruce Lowekamp, Loukas Kallivokas and David O'Hallaron. "*Preliminary Report on the Design of a Framework for Distributed Visualization*". Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA'99), June 1999. pp. 1833–1839.

Rabih Zbib, Saadeddine Mneimneh, Julio López, V. Judson Harward and Richard Rabbat. "*The TRIERARCH Trigger Architecture*". In Proceedings of the Third Annual Federated Laboratory Symposium on Advanced Telecommunications / Information Distribution Research Program, February 1999, pp. 267–271.

Judson Harward, Julio López, Saadeddine Mneimneh, Richard Rabbat, Rabih Zbib. "*An improved Hierarchical Caching Architecture for Low Bandwidth Networks*". In ARL Federated Laboratory Symposium. 1998.

Teaching Experience

Carnegie Mellon University, Pittsburgh, PA
- Teaching assistant, 15-441 Computer Network. Fall 2002
- Teaching assistant, 18-845 Internet services. Spring 2001.
- Teaching assistant, Principles of the Internet. Summer 1999.

Universidad EAFIT, Medellín, Colombia.
- Adjunct Instructor for the Introduction to Computer Networks course. July 1996 - June 1997.
- Teaching assistant, Algorithms. Spring 1993.

Honors

- Winner, along with the CMU Quake group, of the Analytics Challenge at the 2006 Supercomputing Conference (SC). Every year, SC holds a series of competitions in the following categories: Analytics Challenge, Storage Challenge and Bandwidth Challenge. The primary goal of the Analytics Challenge is to provide a forum for researchers, engineers, and analysts to showcase computationally intensive applications that solve real-world, complex problems through the use of rigorous and sophisticated methods of data analysis and high end visualization in conjunction with the use of high performance computing, networking, and storage. Our group won the Analytics Challenge for our work titled "Remote Rutime Steering of Integrated Terascale Simulation and Visualization".

- Winner 2003 Gordon Bell Award for Special Achievement along with the members of the CMU's Quake project. The milestone calculations for the award included:
  - The generation of a record unstructured hex mesh (3.7 billion elements, 4 billion nodes)
  - The largest unstructured mesh wave propagation simulation (900 million elements, 3.2 billion DOF)
  - The largest acoustic wave propagation inverse problem (17 million inversion parameters, 70 billion total unknowns)
  - The largest elastic wave propagation inverse problem (275,000 inversion parameters, something like a billion total unknowns).

- Two times receipient of the scholarship granted by Universidad EAFIT to the highest GPA for the term among all the students in the program, in 1991 and 1992.

- First place in the Mathematics Contest for High School Students organized by Universidad EAFIT in 1990.

- Finalist in the 4th National Mathematics Contest in 1989.

- Third place in the 2nd and 3rd State-wide Mathematics Contest in 1985 and 1987.

Invited Talks and Panels

- Panel on MapReduce and DBMS technologies at SSDBM 2010 in Heidelberg, Germany.

- Tutorial on Data-Intensive Applications, Mass Storage Systems and Technologies Conference, September 2008.

- The case for Data-Intensive Scalable Computing for Science, Los Alamos National Laboratory, July 2008.

Seminars and Colloquia

Data-Intensive Scalable Computing for Science, NASA Goddard's Information Science and Technology (IS&T) Colloquium Series, February 2009.

Software Artifacts

- Principal developer of the CMU Quake Ground Motion Numerical Simulation. Quake is a highly-scalable, parallel FEM numerical solver for simulating large magnitude earthquakes. It is used by the CMU Quake group and seismologists working with the Southern California Earthquake Center (SCEC) to carry out the largest open-science ground motion simulations in peta-scale super computers. This software was instrumental for the prestigious Super Computing (SC) Conference awards we obtained in 2003 and 2006. More information about this software package can be found at the CMU Quake website [`http:/www.cs.cmu.edu/~quake`].

  I am the principal maintainer in charge of adding new features and improving the overall scalability of the code to execute in peta-scale platforms such as the Kraken supercomputer at NICS. During the initial development phase, I designed and developed the mechanisms and module for the parallel output, which is indispensable to write out the results when a simulation executes on a large scale computer with a parallel file systems attached to it. I also made significant contributions to other modules during the initial development including, among others, the module for querying the material models, earthquake source generation, and post-processing query tools for the simulation output.

- BEM Compression and Microsolver query engine for 4D seismic wavefields. This software artifact comprises a set of tools and a library for compressing, querying and processing massive seismic wavefields generated by large-scale ground motion simulations. The main goal is to facilitate the analysis of large wavefields by turning an I/O intensive problem into a massively parallel computation process. It achieves this through clever domain-specific compression schemes and numerical transformations based on the boundary element method (BEM). This approach has also applications to the simulation (wavefield generation) phase by reducing the data size up to a 3:1 ratio through compression and thus lowering the I/O bandwidth requirement at simulation time. A post-simulation step can further compress the data up to a 15:1 ratio. Relevant parts of this library will be integrated with the Quake ground motion simulator to realize the aforementioned reductions in I/O requirement at simulation time. I am the single developer of this software artifact.

- I co-developed the Etree software. The Etree software, available at `http://www.cs.cmu.edu/~euclid`, comprises a library, a set of storage formats and accompanying tools for operating on large out-of-core octrees. The CMU Quake project uses etrees to represent ground models, simulation meshes and indices for 4D wavefields produced by the simulations. Outside CMU, the etree library has been widely used in the seismology community to store and represent models of the ground used in simulation. For example, the US Geological Survey (USGS) uses etrees to represent ground models for the Central California area. These models were used by a group of USGS seismologists to simulate with supercomputers the Great San Francisco Earthquake from 1906 as part of the 100 year commemoration of the event. Similarly, the Southern California Earthquake Center (SCEC) has adopted etrees as the standard representation and API for ground models of Southern California. SCEC is a large, NSF-funded,

collaborative effort that brings together top researchers from seismology, earthquake engineering, civil engineering and computer science.

Programming Languages and Skills

I implemented the parallel I/O system for the CMU Quake ground-motion simulation numerical solver. This software artifact is written in C/C++ and runs on super computers with thousands of processors.

Programming Languages: C/C++, Java, Python, Perl, SQL, Shell (`sh`/`bash`).

Programming Platforms: Hadoop, HBase, Linux, Solaris, Mac OS X, Cray XT5, Cray XT3 Catamount OS, Digital OSF, Windows.

Numerical and Matrix libraries: Message Passing Interface (MPI), POSIX Threads, BLAS / LAPACK / ScaLAPACK, GNU Scientific Library (GSL), Automatically Tuned Linear Algebra Subroutines (ATLAS), VxL, Visualization Toolkit (VTK), Insight Registration and Segmentation Toolkit (ITK).

Service and Volunteer Work

Program committee member for:
- Workshop on Array Databases 2011.
- Storage track of the IEEE international conference on Networking, Architecture, and Storage (NAS 2011).

From October 2004 to August 2007, I was in charge of setting up, maintaining and administering the computing and storage infrastructure for the CMU Quake and Computational Databases System projects.

In May 2006, along with two other CMU graduate students, created a tennis club for the campus community. This volunteer-run organization brings together more than 150 players from the campus community and coordinates competition and outreach events.

In the Spring of 1999 I served as volunteer in events organized by the CMU Office of International Education (OIE) in coordination with the Internal Revenue Service (IRS). The goal of these events was to provide tax-related information and guidance to foreign members of the campus community.

Additional Information

Languages: Proficient English; native Spanish.
Professional Affiliations: Institute of Electrical and Electronics Engineers (IEEE), Association of Computing Machinery (ACM).