The Story of GraphLab – From Scaling Machine Learning to Shaping Graph Systems Research (VLDB 2023 Test-of-time Award Talk)

Joseph E. Gonzalez
UC Berkeley
jegonzal@berkeley.edu

Yucheng Low
XetHub Inc.
ylow@xethub.com

ABSTRACT

The GraphLab project spanned almost a decade and had profound academic and industrial impact on large-scale machine learning and graph processing systems. There were numerous papers written describing the innovations in GraphLab including the original vertex-centric [8] and edge-centric [3] programming abstractions, high-performance asynchronous execution engines [9], out-of-core graph computation [6], tabular graph-systems [4], and even new statistical inference algorithms [2] enabled by the GraphLab project. This work became the basis of multiple PhD theses [1, 5, 7]. The GraphLab open-source project had broad academic and industrial adoption and ultimately lead to the launch of Turi.

In this talk, we tell the story of GraphLab, how it began and the key ideas behind it. We will focus on the approach to achieving scalable asynchronous systems in machine learning. During our talk, we will explore the impact that GraphLab has had on the development of graph processing systems, graph databases, and AI/ML; Additionally, we will share our insights and opinions into where we see the future of these fields heading. In the process, we highlight some of the lessons we learned and provide guidance for future students.

PVLDB Reference Format:
Joseph E. Gonzalez and Yucheng Low. The Story of GraphLab – From Scaling Machine Learning to Shaping Graph Systems Research (VLDB 2023 Test-of-time Award Talk). PVLDB, 16(12): 4138 - 4138, 2023.
doi:10.14778/3611540.3611637

SPEAKER BIOGRAPHIES

Joseph E. Gonzalez

Joseph Gonzalez is an Associate Professor of Computer Science at UC Berkeley. He is the co-director of the RISE and Sky Computing Labs and a member of the Berkeley AI Research (BAIR) group. Joseph is also co-founder and VP of product at Aqueduct. His research addresses problems in data systems, neural network design, compilers and distributed systems for large scale machine learning, natural language processing, computer vision, robotics, autonomous driving, and graph analytics. Prior to joining Berkeley, Gonzalez co-founded Turi Inc (formerly GraphLab and Dato) based on his thesis work and created the GraphX project (now part of Apache Spark). Gonzalez's innovative work has earned him significant recognition, including the Okawa Research Grant, the NSF Expedition Award, and the NSF Early CAREER Award.

Yucheng Low

Yucheng Low is the CEO and co-founder of Xethub, building the next generation ML data management platform. Yucheng's PhD thesis at Carnegie Mellon University focused on large-scale Machine Learning with the GraphLab abstraction. In 2013, based on his thesis work, he co-founded Turi (formerly known as GraphLab and Dato), where he was the Chief Architect and played a crucial role in developing scalable single machine DataFrame libraries and ML algorithms. Turi was acquired by Apple in 2016. At Apple, Yucheng contributed to broad aspects of the internal ML platform stack, spanning storage to inference.

ACKNOWLEDGMENTS

The GraphLab project was a team effort made possible by our thesis advisor Carlos Guestrin and our close collaborators and colleagues: Danny Bickson, Haijie Gu, Joseph M. Hellerstein, and Aapo Kyrola.

REFERENCES

[1] Joseph Gonzalez. 2012. *Parallel and Distributed Systems for Probabilistic Reasoning*. Ph.D. Dissertation. USA. Advisor(s) Guestrin, Carlos. AAI3538976.
[2] Joseph E. Gonzalez, Yucheng Low, Arthur Gretton, and Carlos Guestrin. 2011. Parallel Gibbs Sampling: From Colored Fields to Thin Junction Trees. In *Artificial Intelligence and Statistics (AISTATS)*. http://proceedings.mlr.press/v15/gonzalez11a.html
[3] Joseph E. Gonzalez, Yucheng Low, Haijie Gu, Danny Bickson, and Carlos Guestrin. 2012. PowerGraph: Distributed Graph-Parallel Computation on Natural Graphs. In *OSDI '12*. https://www.usenix.org/system/files/conference/osdi12/osdi12-final-167.pdf
[4] Joseph E. Gonzalez, Reynold S. Xin, Ankur Dave, Daniel Crankshaw, Michael J. Franklin, and Ion Stoica. 2014. GraphX: Graph Processing in a Distributed Dataflow Framework. In *11th USENIX Symposium on Operating Systems Design and Implementation (OSDI 14)*. 599–613.
[5] Aapo Kyrola. 2014. *Large-scale Graph Computation on Just a PC*. Ph.D. Dissertation. USA. Advisor(s) Guestrin, Carlos.
[6] Aapo Kyrola, Guy Blelloch, and Carlos Guestrin. 2012. GraphChi: Large-Scale Graph Computation on Just a PC. In *10th USENIX Symposium on Operating Systems Design and Implementation (OSDI 12)*. USENIX Association, Hollywood, CA, 31–46. https://www.usenix.org/conference/osdi12/technical-sessions/presentation/kyrola
[7] Yucheng Low. 2013. *GraphLab: A Distributed Abstraction for Large Scale Machine Learning*. Ph.D. Dissertation. USA. Advisor(s) Guestrin, Carlos.
[8] Yucheng Low, Joseph E. Gonzalez, Aapo Kyrola, Daniel Bickson, Carlos Guestrin, and Joseph M. Hellerstein. 2010. GraphLab: A New Parallel Framework for Machine Learning. In *Conference on Uncertainty in Artificial Intelligence (UAI)*. https://arxiv.org/abs/1006.4990
[9] Yucheng Low, Joseph E. Gonzalez, Aapo Kyrola, Danny Bickson, Carlos Guestrin, and Joseph M. Hellerstein. 2012. Distributed GraphLab: A Framework for Machine Learning and Data Mining in the Cloud.. In *Proceedings of Very Large Data Bases (PVLDB)*. https://arxiv.org/abs/1204.6078

Proceedings of the VLDB Endowment, Vol. 16, No. 12 ISSN 2150-8097.
doi:10.14778/3611540.3611637