Understanding the Network Performance of Graph Processing Frameworks

Anil Yelam, Audrey Randall

Abstract

An abstract is like a movie trailer. It offers a preview, highlights key points, and helps the audience decide whether to view the entire work. Abstracts are the pivot of a research paper because many journal editorial boards screen manuscripts only on the basis of the abstract.

1 Introduction

The importance of balanced execution of big data frameworks (Papers to link: Osterhout[8], Themis[9], TritonSort[10]).

Why we picked graph processing.

What we did or found out, in a nutshell.

2 Background

All about the graph processing frameworks and their computational models, basically a brief summary of [6]. And a bunch of related papers we have read, like comparison papers [5, 1] and some earlier graph processing systems these papers point to - such as Pregel[7], PowerGraph [4] and others.

3 Apache Giraph

Description of Apache Giraph[3]. And bunch of optimizations made by Facebook[2].

4 Implementation

All that we did! (be super verbose.)

References

[1] AMMAR, K., AND ÖZSU, M. T. Experimental analysis of distributed graph systems. *Proc. VLDB Endow. 11*, 10 (June 2018), 1151–1164.

[2] CHING, A., EDUNOV, S., KABILJO, M., LOGOTHETIS, D., AND MUTHUKRISHNAN, S. One trillion edges: Graph processing at facebook-scale. *Proc. VLDB Endow. 8*, 12 (Aug. 2015), 1804–1815.

[3] GIRAPH. http://giraph.apache.org/. accessed: 2019-05-03.

[4] GONZALEZ, J. E., LOW, Y., GU, H., BICKSON, D., AND GUESTRIN, C. Powergraph: Distributed graph-parallel computation on natural graphs. In *Proceedings of the 10th USENIX Conference on Operating Systems Design and Implementation* (Berkeley, CA, USA, 2012), OSDI'12, USENIX Association, pp. 17–30.

[5] GUO, Y., BICZAK, M., VARBANESCU, A. L., IOSUP, A., MARTELLA, C., AND WILLKE, T. L. How well do graph-processing platforms perform? an empirical performance evaluation and analysis. In *Proceedings of the 2014 IEEE 28th International Parallel and Distributed Processing Symposium* (Washington, DC, USA, 2014), IPDPS '14, IEEE Computer Society, pp. 395–404.

[6] HEIDARI, S., SIMMHAN, Y., CALHEIROS, R. N., AND BUYYA, R. Scalable graph processing frameworks: A taxonomy and open challenges. *ACM Comput. Surv. 51*, 3 (June 2018), 60:1–60:53.

[7] MALEWICZ, G., AUSTERN, M. H., BIK, A. J., DEHNERT, J. C., HORN, I., LEISER, N., AND CZAJKOWSKI, G. Pregel: A system for large-scale graph processing. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data* (New York, NY, USA, 2010), SIGMOD '10, ACM, pp. 135–146.

[8] OUSTERHOUT, K., RASTI, R., RATNASAMY, S., SHENKER, S., AND CHUN, B.-G. Making sense of performance in data analytics frameworks. In *Proceedings of the 12th USENIX Conference on Networked Systems Design and Implementation* (Berkeley, CA, USA, 2015), NSDI'15, USENIX Association, pp. 293–307.

[9] RASMUSSEN, A., LAM, V. T., CONLEY, M., PORTER, G., KAPOOR, R., AND VAHDAT, A. Themis: An i/o-efficient mapreduce. In *Proceedings of the Third ACM Symposium on Cloud Computing* (New York, NY, USA, 2012), SoCC '12, ACM, pp. 13:1–13:14.

[10] RASMUSSEN, A., PORTER, G., CONLEY, M., MADHYASTHA, H. V., MYSORE, R. N., PUCHER, A., AND VAHDAT, A. Tritonsort: A balanced and energy-efficient large-scale sorting system. *ACM Trans. Comput. Syst. 31*, 1 (Feb. 2013), 3:1–3:28.