the clouds: A Berkeley view of cloud computing. *UC Berkeley Technical Report UCB/EECS-2009-28*, February 2009.

[8] P. Barham, B. Dragovic, K. Fraser, S. Hand, T. Harris, A. Ho, R. Neugebauer, I. Pratt, and A. Warfield. Xen and the art of virtualization. In *Proceedings of the 19th ACM symposium on Operating Systems Principles*, pages 164–177, 2003.

[9] L. Barroso and U. Hölzle. The datacenter as a computer: An introduction to the design of warehouse-scale machines. *Synthesis Lectures on Computer Architecture*, 4(1):1–108, 2009.

[10] L. Barroso, J. Dean, and U. Hoelzle. Web search for a planet: The Google cluster architecture. *IEEE Micro*, 23(2):22–28, 2003.

[11] K. Birman, G. Chockler, and R. van Renesse. Toward a cloud computing research agenda. *SIGACT News*, 40(2):68–80, 2009.

[12] R. Buyya, T. Cortes, and H. Jin. Single system image. *Intl. Journal of High Performance Computing Applications*, 15(2):124, 2001.

[13] B. Chamberlain, D. Callahan, and H. Zima. Parallel programmability and the Chapel language. *International Journal of High Performance Computing Applications*, 21(3):291, 2007.

[14] M. Chapman and G. Heiser. vNUMA: A virtual shared-memory multiprocessor. In *Proceedings of the 2009 conference on USENIX Annual technical conference*, 2009.

[15] P. Charles, C. Grothoff, V. Saraswat, C. Donawa, A. Kielstra, K. Ebcioglu, C. Von Praun, and V. Sarkar. X10: an object-oriented approach to non-uniform cluster computing. In *ACM SIGPLAN Notices*, volume 40, pages 519–538, 2005.

[16] D.-K. Chen, H.-M. Su, and P.-C. Yew. The impact of synchronization and granularity on parallel systems. In *Proceedings of the 17th annual intl. symposium on Computer Architecture*, pages 239–248, 1990.

[17] Y. Chen, D. Pavlov, and J. F. Canny. Large-scale behavioral targeting. In *Proc. of the 15th ACM SIGKDD intl conf. on Knowledge discovery and data mining*, pages 209–218, 2009.

[18] C.-T. Chu, S. K. Kim, Y.-A. Lin, Y. Yu, G. Bradski, A. Y. Ng, and K. Olukotun. Map-Reduce for machine learning on multicore. In *Proc. of NIPS'07*, pages 281–288, 2007.

[19] T. Condie, N. Conway, P. Alvaro, J. Hellerstein, K. Elmeleegy, and R. Sears. MapReduce online. In *Proceedings of the 7th USENIX conf. on networked systems design and implementation*, pages 21–21, 2010.

[20] J. Dean and S. Ghemawat. MapReduce: simplified data processing on large clusters. In *the 6th Conference on Symposium on Operating Systems Design & Implementation*, volume 6, pages 137–150, 2004.

[21] J. Ekanayake, S. Pallickara, and G. Fox. MapReduce for data intensive scientific analysis. In *Fourth IEEE International Conference on eScience*, pages 277–284, 2008.

[22] M. P. I. Forum. MPI: A message-passing interface standard. `http://www.mpi-forum.org/docs/mpi-2.2/mpi22-report.pdf`, 2009. [last access: 11/2, 2011].

[23] S. Ghemawat, H. Gobioff, and S.-T. Leung. The Google file system. In *Proc. of the 9th ACM Symposium on Operating Systems Principles (SOSP'03)*, pages 29–43, 2003.

[24] B. Hayes. Cloud computing. *Communications of the ACM*, 51(7): 9–11, 2008.

[25] B. He, W. Fang, Q. Luo, N. Govindaraju, and T. Wang. Mars: a MapReduce framework on graphics processors. In *Proceedings of the 17th international conference on parallel architectures and compilation techniques*, pages 260–269, 2008.

[26] B. Hedlund. Inverse virtualization for internet scale applications. `http://bradhedlund.com/2011/03/16/inverse-virtualization-for-internet-scale-applications/`. [last access: 11/2, 2011].

[27] M. Isard, M. Budiu, Y. Yu, A. Birrell, and D. Fetterly. Dryad: distributed data-parallel programs from sequential building blocks. In *EuroSys '07: Proceedings of the 2nd ACM SIGOPS/EuroSys European Conference on Computer Systems 2007*, pages 59–72, 2007.

[28] H. Jégou, M. Douze, and C. Schmid. Improving bag-of-features for large scale image search. *International Journal of Computer Vision*, 87(3):316–336, 2010.

[29] P. Keleher, A. Cox, S. Dwarkadas, and W. Treadmarks. Distributed shared memory on standard workstations and operating systems. In *Proc. 1994 Winter Usenix Conference*, pages 115–131, 1994.

[30] A. Kivity, Y. Kamay, D. Laor, U. Lublin, and A. Liguori. KVM: the linux virtual machine monitor. In *Proceedings of the Linux Symposium*, volume 1, pages 225–230, 2007.

[31] D. Lee, S. Baek, and K. Sung. Modified k-means algorithm for vector quantizer design. *Signal Processing Letters, IEEE*, 4(1):2–4, 1997.

[32] K. Li and P. Hudak. Memory coherence in shared virtual memory systems. *ACM Trans. Comput. Syst.*, 7(4):321–359, 1989.

[33] H. Lu, S. Dwarkadas, A. Cox, and W. Zwaenepoel. Message passing versus distributed shared memory on networks of workstations. In *Proc. of the IEEE/ACM Supercomputing 95 Conf.*, page 37, 1995.

[34] Z. Ma and L. Gu. The limitation of MapReduce: A probing case and a lightweight solution. In *Proc. of the 1st Intl. Conf. on Cloud Computing, GRIDs, and Virtualization*, pages 68–73, 2010.

[35] J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, page 14, 1967.

[36] MathWorks. Inc. Matlab. `http://www.mathworks.com/products/matlab/`. [last access: 11/2, 2011].

[37] B. Nitzberg and V. Lo. Distributed shared memory: A survey of issues and algorithms. *Computer*, 24(8):52–60, 1991.

[38] D. Nurmi, R. Wolski, C. Grzegorczyk, G. Obertelli, S. Soman, L. Youseff, and D. Zagorodnov. The Eucalyptus open-source cloud-computing system. In *Proc. of the 9th IEEE/ACM Intl. Symposium on Cluster Computing and the Grid*, pages 124–131, 2009.

[39] P. J. Nurnberg, U. K. Wiil, and D. L. Hicks. A grand unified theory for structural computing. *Metainformatics*, 3002:1–16, 2004.

[40] R. Pike, S. Dorward, R. Griesemer, and S. Quinlan. Interpreting the data: Parallel analysis with Sawzall. *Sci. Program.*, 13(4):277–298, 2005.

[41] C. Ranger, R. Raghuraman, A. Penmetsa, G. Bradski, and C. Kozyrakis. Evaluating MapReduce for multi-core and multiprocessor systems. In *Proc. of the 2007 IEEE 13th Intl Symposium on High Performance Computer Architecture*, pages 13–24, 2007.

[42] Salesforce.com. `http://www.salesforce.com`. [last access: 11/2, 2011].

[43] M. C. Schatz. CloudBurst: highly sensitive read mapping with MapReduce. *Bioinformatics*, 25:1363–1369, 2009.

[44] The R Project. The R Language. `http://www.r-project.org/`. [last access: 11/2, 2011].

[45] C. Tseng. Compiler optimizations for eliminating barrier synchronization. In *ACM SIGPLAN Notices*, volume 30, pages 144–155, 1995.

[46] C. A. Waldspurger. Memory resource management in VMware ESX server. *SIGOPS Oper. Syst. Rev.*, 36(SI):181–194, 2002.

[47] H.-C. Yang, A. Dasdan, R.-L. Hsiao, and D. S. Parker. Map-Reduce-Merge: simplified relational data processing on large clusters. In *SIGMOD '07: Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, pages 1029–1040, 2007.

[48] Y. Yu, M. Isard, D. Fetterly, M. Budiu, Ú. Erlingsson, P. K. Gunda, and J. Currey. DryadLINQ: A system for general-purpose distributed data-parallel computing using a high-level language. In *the 8th Conference on Symposium on Operating Systems Design & Implementation*, pages 1–14, 2008.

[49] M. Zaharia, D. Borthakur, J. Sen Sarma, K. Elmeleegy, S. Shenker, and I. Stoica. Delay scheduling: a simple technique for achieving locality and fairness in cluster scheduling. In *EuroSys '10: Proceedings of the 5th European conference on computer systems*, pages 265–278, 2010.

[50] R. Zhang and A. Rudnicky. A large scale clustering scheme for kernel k-means. *Pattern Recognition*, 4:40289, 2002.

[51] W. Zhao, H. Ma, and Q. He. Parallel k-means clustering based on mapreduce. In *roceedings of the First International Conference on Cloud Computiong (CloudCom)*, pages 674–679, 2009.