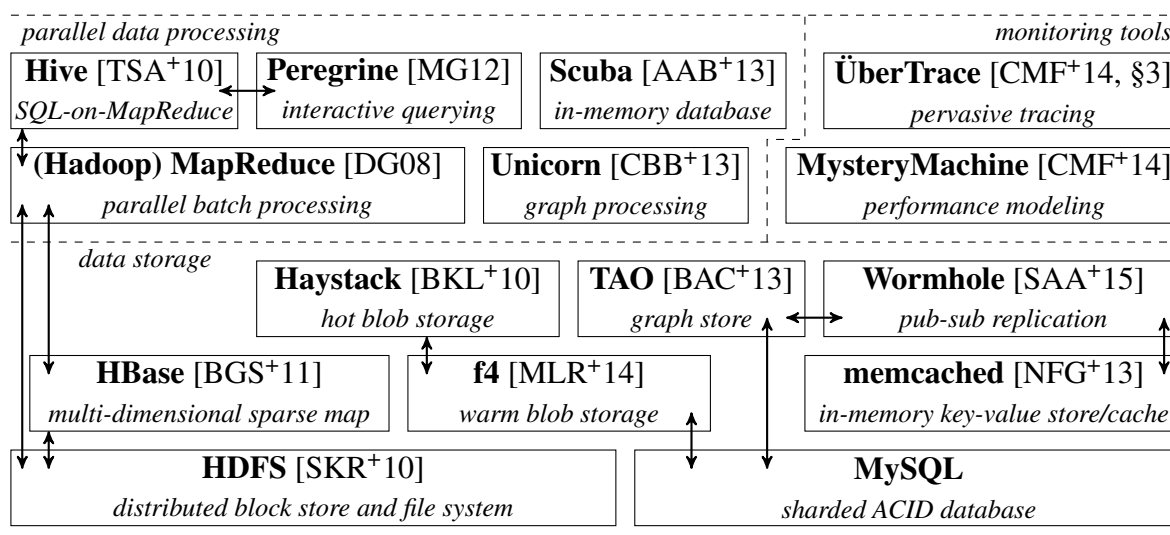


## The Facebook Stack

**Source:** Malte Schwarzkopf. “Operating system support for warehouse-scale computing”. PhD thesis. University of Cambridge Computer Laboratory (to appear), 2015, Chapter 2.



**Figure 1:** The Facebook infrastructure stack. I omit front-end serving systems about which details are unknown. Arrows indicate data exchange and dependencies between systems; simple layering does *not* imply a dependency or relation.

In addition, there are several papers that do not directly cover systems in the Facebook stack, but describe workloads, techniques or data centre hardware:

- Descriptions of the physical design of Facebook’s server machines as of 2011 [FHL<sup>+</sup>11] and data centre network architecture as of 2013 [FA13].
- Another paper on the HBase back-end for Facebook messages [ABC<sup>+</sup>12] and a measurement paper looking at the HDFS-level usage patterns of this HBase deployment [HBD<sup>+</sup>14].
- Papers on the use of erasure codes in HDFS at Facebook [RSG<sup>+</sup>13; SAP<sup>+</sup>13; RSG<sup>+</sup>14].
- Several papers analysing the Facebook memcached workload [AXF<sup>+</sup>12] and evaluating new sampling strategies to improve hit rates in memcached [LLD<sup>+</sup>13].
- A study of Facebook’s wide-area photo caching infrastructure [HBR<sup>+</sup>13].
- A description of how Facebook uses shared memory to persist in-memory state across restarts of Scuba server processes [GCG<sup>+</sup>14].
- The HipHop Virtual Machine (HHVM) is a JIT compiler and runtime for PHP code heavily used in front-end page generation [AEM<sup>+</sup>14]. Previously, Facebook used a source-to-source compiler (also called “HipHop”, HPHPC) to transform PHP into semantically equivalent C++ code that can be compiled into native code [ZPY<sup>+</sup>12].

# Bibliography

- [AAB<sup>+</sup>13] Lior Abraham, John Allen, Oleksandr Barykin, Vinayak Borkar, Bhuwan Chopra, Ciprian Gerea, et al. “Scuba: Diving into Data at Facebook”. In: *Proceedings of the VLDB Endowment* 6.11 (Aug. 2013), pp. 1057–1067 (cited on page 1).
- [AEM<sup>+</sup>14] Keith Adams, Jason Evans, Bertrand Maher, Guilherme Ottoni, Andrew Paroski, Brett Simmers, et al. “The Hiphop Virtual Machine”. In: *Proceedings of the 2014 ACM International Conference on Object Oriented Programming Systems Languages and Applications (OOPSLA)*. Portland, Oregon, USA, 2014, pp. 777–790 (cited on page 1).
- [ABC<sup>+</sup>12] Amitanand S Aiyer, Mikhail Bautin, Guoqiang Jerry Chen, Pritam Damania, Prakash Khemani, Kannan Muthukkaruppan, et al. “Storage Infrastructure Behind Facebook Messages: Using HBase at Scale.” In: *IEEE Data Engineering Bulletin* 35.2 (2012), pp. 4–13 (cited on page 1).
- [AXF<sup>+</sup>12] Berk Atikoglu, Yuehai Xu, Eitan Frachtenberg, Song Jiang, and Mike Paleczny. “Workload Analysis of a Large-scale Key-value Store”. In: *Proceedings of the 12<sup>th</sup> ACM SIGMETRICS/PERFORMANCE Joint International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS)*. London, England, UK, 2012, pp. 53–64 (cited on page 1).
- [BKL<sup>+</sup>10] Doug Beaver, Sanjeev Kumar, Harry C. Li, Jason Sobel, and Peter Vajgel. “Finding a Needle in Haystack: Facebook’s Photo Storage”. In: *Proceedings of the 9<sup>th</sup> USENIX Conference on Operating Systems Design and Implementation (OSDI)*. Vancouver, British Columbia, Canada, 2010, pp. 1–8 (cited on page 1).
- [BGS<sup>+</sup>11] Dhruva Borthakur, Jonathan Gray, Joydeep Sen Sarma, Kannan Muthukkaruppan, Nicolas Spiegelberg, Hairong Kuang, et al. “Apache Hadoop Goes Realtime at Facebook”. In: *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data (SIGMOD)*. Athens, Greece, 2011, pp. 1071–1080 (cited on page 1).
- [BAC<sup>+</sup>13] Nathan Bronson, Zach Amsden, George Cabrera, Prasad Chakka, Peter Dimov, Hui Ding, et al. “TAO: Facebook’s Distributed Data Store for the Social Graph”. In: *Proceedings of the 2013 USENIX Annual Technical Conference*. San Jose, California, USA, 2013, pp. 49–60 (cited on page 1).

- [CMF<sup>+</sup>14] Michael Chow, David Meisner, Jason Flinn, Daniel Peek, and Thomas F. Wenisch. “The Mystery Machine: End-to-end Performance Analysis of Large-scale Internet Services”. In: *Proceedings of the 11<sup>th</sup> USENIX Conference on Operating Systems Design and Implementation (OSDI)*. Broomfield, Colorado, USA, 2014, pp. 217–231 (cited on page 1).
- [CBB<sup>+</sup>13] Michael Curtiss, Iain Becker, Tudor Bosman, Sergey Doroshenko, Lucian Gri-jincu, Tom Jackson, et al. “Unicorn: A System for Searching the Social Graph”. In: *Proceedings of the VLDB Endowment* 6.11 (Aug. 2013), pp. 1150–1161 (cited on page 1).
- [DG08] Jeffrey Dean and Sanjay Ghemawat. “MapReduce: Simplified Data Processing on Large Clusters”. In: *Communications of the ACM* 51.1 (Jan. 2008), pp. 107–113 (cited on page 1).
- [FA13] Nathan Farrington and Alexey Andreyev. “Facebook’s data center network architecture”. In: *Proceedings of the 2013 IEEE Optical Interconnects Conference*. May 2013, pp. 49–50 (cited on page 1).
- [FHL<sup>+</sup>11] Eitan Frachtenberg, Ali Heydari, Harry Li, Amir Michael, Jacob Na, Avery Nisbet, et al. “High-efficiency Server Design”. In: *Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis (SC)*. Seattle, Washington, USA, 2011, 27:1–27:11 (cited on page 1).
- [GCG<sup>+</sup>14] Aakash Goel, Bhuwan Chopra, Ciprian Gerea, Dhruv Mátáni, Josh Metzler, Fahim Ul Haq, et al. “Fast Database Restarts at Facebook”. In: *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data (SIGMOD)*. Snowbird, Utah, USA, 2014, pp. 541–549 (cited on page 1).
- [HBD<sup>+</sup>14] Tyler Harter, Dhruva Borthakur, Siying Dong, Amitanand Aiyer, Liyin Tang, Andrea C. Arpaci-Dusseau, et al. “Analysis of HDFS Under HBase: A Facebook Messages Case Study”. In: *Proceedings of the 12<sup>th</sup> USENIX Conference on File and Storage Technologies (FAST)*. Santa Clara, California, USA, 2014, pp. 199–212 (cited on page 1).
- [HBR<sup>+</sup>13] Qi Huang, Ken Birman, Robbert van Renesse, Wyatt Lloyd, Sanjeev Kumar, and Harry C. Li. “An Analysis of Facebook Photo Caching”. In: *Proceedings of the 24<sup>th</sup> ACM Symposium on Operating Systems Principles (SOSP)*. Farmington, Pennsylvania, USA, 2013, pp. 167–181 (cited on page 1).
- [LLD<sup>+</sup>13] John Liang, James Luo, Mark Drayton, Rajesh Nishtala, Richard Liu, Nick Hammer, et al. “Storage and Performance Optimization of Long Tail Key Access in a Social Network”. In: *Proceedings of the 3<sup>rd</sup> International Workshop on Cloud Data and Platforms (CloudDP)*. Prague, Czech Republic, 2013, pp. 1–6 (cited on page 1).

- [MLR<sup>+</sup>14] Subramanian Muralidhar, Wyatt Lloyd, Sabyasachi Roy, Cory Hill, Ernest Lin, Weiwen Liu, et al. “f4: Facebook’s Warm BLOB Storage System”. In: *Proceedings of the 11<sup>th</sup> USENIX Conference on Operating Systems Design and Implementation (OSDI)*. OSDI’14. Broomfield, Colorado, USA, 2014, pp. 383–398 (cited on page 1).
- [MG12] Raghotham Murthy and Rajat Goel. “Peregrine: Low-latency Queries on Hive Warehouse Data”. In: *XRDS: Crossroad, ACM Magazine for Students* 19.1 (Sept. 2012), pp. 40–43 (cited on page 1).
- [NFG<sup>+</sup>13] Rajesh Nishtala, Hans Fugal, Steven Grimm, Marc Kwiatkowski, Herman Lee, Harry C. Li, et al. “Scaling Memcache at Facebook”. In: *Proceedings of the 10<sup>th</sup> USENIX Conference on Networked Systems Design and Implementation (NSDI)*. Lombard, IL, 2013, pp. 385–398 (cited on page 1).
- [RSG<sup>+</sup>13] K. V. Rashmi, Nihar B. Shah, Dikang Gu, Hairong Kuang, Dhruva Borthakur, and Kannan Ramchandran. “A Solution to the Network Challenges of Data Recovery in Erasure-coded Distributed Storage Systems: A Study on the Facebook Warehouse Cluster”. In: *Proceedings of the 5<sup>th</sup> USENIX Workshop on Hot Topics in Storage and File Systems (HotStorage)*. San Jose, CA, 2013, pp. 8–8 (cited on page 1).
- [RSG<sup>+</sup>14] K.V. Rashmi, Nihar B. Shah, Dikang Gu, Hairong Kuang, Dhruva Borthakur, and Kannan Ramchandran. “A “Hitchhiker’s” Guide to Fast and Efficient Data Reconstruction in Erasure-coded Data Centers”. In: *Proceedings of the 2014 ACM SIGCOMM Conference (SIGCOMM)*. Chicago, Illinois, USA, 2014, pp. 331–342 (cited on page 1).
- [SAP<sup>+</sup>13] Maheswaran Sathiamoorthy, Megasthenis Asteris, Dimitris Papailiopoulos, Alexandros G. Dimakis, Ramkumar Vadali, Scott Chen, et al. “XORing elephants: novel erasure codes for big data”. In: *Proceedings of the 39<sup>th</sup> International Conference on Very Large Data Bases (VLDB)*. Trento, Italy, 2013, pp. 325–336 (cited on page 1).
- [Sch15] Malte Schwarzkopf. “Operating system support for warehouse-scale computing”. PhD thesis. University of Cambridge Computer Laboratory (to appear), 2015 (cited on page 1).
- [SAA<sup>+</sup>15] Yogeshwer Sharma, Philippe Ajoux, Petchean Ang, David Callies, Abhishek Choudhary, Laurent Demailly, et al. “Wormhole: Reliable Pub-Sub to Support Georeplicated Internet Services”. In: *Proceedings of the 12<sup>th</sup> USENIX Symposium on Networked Systems Design and Implementation (NSDI)*. Oakland, CA, May 2015, pp. 351–366 (cited on page 1).

- 
- [SKR<sup>+</sup>10] Konstantin Shvachko, Hairong Kuang, Sanjay Radia, and Robert Chansler. “The Hadoop Distributed File System”. In: *Proceedings of the 26<sup>th</sup> IEEE Symposium on Mass Storage Systems and Technologies (MSST)*. May 2010, pp. 1–10 (cited on page 1).
- [TSA<sup>+</sup>10] Ashish Thusoo, Zheng Shao, Suresh Anthony, Dhruva Borthakur, Namit Jain, Joydeep Sen Sarma, et al. “Data Warehousing and Analytics Infrastructure at Facebook”. In: *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data (SIGMOD)*. Indianapolis, Indiana, USA, 2010, pp. 1013–1020 (cited on page 1).
- [ZPY<sup>+</sup>12] Haiping Zhao, Iain Proctor, Minghui Yang, Xin Qi, Mark Williams, Qi Gao, et al. “The HipHop Compiler for PHP”. In: *Proceedings of the ACM 2012 International Conference on Object Oriented Programming Systems Languages and Applications (OOPSLA)*. Tucson, Arizona, USA, 2012, pp. 575–586 (cited on page 1).