

0xdata. *H2O*, 10:27:59. <https://github.com/0xdata/h2o>.

3 Round Stones. *Callimachus*, 21:58:48. <http://callimachusproject.org/>.

451 Research. "Database Landscape Map," n.d. [http://blogs.the451group.com/information\\_management/files/2013/02/db\\_Map\\_2\\_13.jpg](http://blogs.the451group.com/information_management/files/2013/02/db_Map_2_13.jpg).

Abadi, Daniel. "Classifying the SQL-on-Hadoop Solutions." *Hadapt*. Accessed October 17, 2013. <http://hadapt.com/blog/2013/10/02/classifying-the-sql-on-hadoop-solutions/>.

Abelson, Harold. *Structure and Interpretation of Computer Programs*. 2nd ed. Cambridge, Mass.: New York: MIT Press ; McGraw-Hill, 1996.

Abouzeid, Azza, Kamil Bajda-Pawlikowski, Daniel J. Abadi, and Avi Silberschatz. *HadoopDB*. Accessed October 17, 2013. <http://sourceforge.net/projects/hadoopdb/>.

Abu-Mostafa, Yaser S, Magdon-Ismail, and Hsuan-Tien Lin. *Learning from Data: a Short Course*. [United States]: AMLBook.com, 2012.

Agarwal, Rohit. "Sparse Matrix Multiplication Using SQL," 23:48:56. <http://notes.mindprince.in/2013/06/07/sparse-matrix-multiplication-using-sql.html>.

Agarwal, Sameer, and AMPLab at UC Berkeley. *BlinkDB*. Accessed October 20, 2013. <https://github.com/sameeragarwal/blinkdb>.

Agarwal, Sameer, Anand P. Iyer, Aurojit Panda, Samuel Madden, Barzan Mozafari, and Ion Stoica. "Blink and It's Done: Interactive Queries on Very Large Data." *Proceedings of the VLDB Endowment* 5, no. 12 (2012): 1902–1905.

AGBeat. "Visualizing the Differences Between UX and UI," 23:24:08. <http://agbeat.com/tech-news/visualizing-the-differences-between-ux-and-ui/>.

Agrawal, Ritesh. "On Writing Python UDF for Pig: A Perspective." *Memento*, 20:29:41. <http://ragrawal.wordpress.com/2013/02/24/on-writing-python-udf-for-pig-a-perspective/>.

Airbnb. *Chronos*, 06:48:36. <https://github.com/airbnb/chronos>.

Alag, Satnam. *Collective Intelligence in Action*. Greenwich, Conn.: Manning, 2008.

AlchemyAPI. *AlchemyAPI*, 12:42:42. <http://www.alchemyapi.com/>.

Alves, Alexandre, Lloyd Williams, and Robin J. Smith. "An Overview of Complex Event Processing," n.d. <http://www.packtpub.com/article/overview-complex-event-processing>.

Amazon Web Services. *Amazon DynamoDB*, 10:25:15. <http://aws.amazon.com/dynamodb/>.

———. *Amazon Elastic MapReduce (EMR)*, n.d. <http://aws.amazon.com/elasticmapreduce/>.

———. *Amazon Elastic MapReduce Developer Guide*, n.d. <http://amzn.com/B007US6CIO>.

———. *Amazon Redshift*. Amazon Web Services, Inc., 12:44:29. <http://aws.amazon.com/redshift/>.

———. *EMR Bootstrap Actions*. Accessed October 17, 2013. <https://github.com/AmazonEMR/bootstrap.actions>.

———. *EMR Sample Apps*. Accessed October 17, 2013. <https://github.com/AmazonEMR/sample.apps>.

AMPLab at UC Berkeley. *GraphX*. Accessed October 20, 2013. <https://github.com/amplab/graphx>.

———. *MLbase*. Accessed October 20, 2013. <http://www.mlbase.org/>.

———. *Shark*, 08:52:11. <https://github.com/amplab/shark>.

———. *Tachyon*. Accessed October 20, 2013. <https://github.com/amplab/tachyon>.

Anand, Sid. "LinkedIn's Segmentation & Targeting Platform." *Business & Mgmt*, June 26, 2013. <http://www.slideshare.net/r39132/linkedins-segmentation-targeting-platform?ref=http://eventifier.co/event/hadoops Summit2013/slides>.

- Anderson, Jesse. "How-to: Use Eclipse with MapReduce in Cloudera's QuickStart VM," n.d.  
<http://blog.cloudera.com/blog/2013/08/how-to-use-eclipse-with-mapreduce-in-clouderas-quickstart-vm/>.
- Anil, Robin, Ted Dunning, and Ellen Friedman. *Mahout in Action*. Manning Publications Company, 2011.
- Apache Software Foundation. *Apache Accumulo*. Accessed October 20, 2013.  
<http://accumulo.apache.org/>.
- . *Apache Ambari*, 21:46:13. <http://incubator.apache.org/ambari/>.
- . *Apache Avro*, 10:14:10. <http://avro.apache.org/>.
- . *Apache Cassandra*, 14:08:44. <http://cassandra.apache.org/>.
- . *Apache Crunch*, 21:17:56. <http://crunch.apache.org/>.
- . *Apache Drill*, 14:23:49. <http://incubator.apache.org/drill/>.
- . *Apache Falcon*, 10:17:04. <http://falcon.incubator.apache.org/>.
- . *Apache Flume*, n.d. <http://flume.apache.org/>.
- . *Apache Giraph*, 21:38:50. <http://giraph.apache.org/>.
- . *Apache Hadoop*, 14:09:18. <http://hadoop.apache.org/>.
- . *Apache Hadoop YARN*, 09:39:12.  
<http://hadoop.apache.org/docs/current/hadoop-yarn/hadoop-yarn-site/YARN.html>.
- . *Apache HBase*, 21:25:16. <http://hbase.apache.org/>.
- . *Apache Hive*, n.d. <http://hive.apache.org/>.
- . *Apache Kafka*, n.d. <http://kafka.apache.org>.
- . *Apache Lucene*, 09:15:57. <http://lucene.apache.org/>.
- . *Apache Mahout*, 09:12:48. <http://mahout.apache.org/>.
- . *Apache Mesos*, n.d. <http://mesos.apache.org/>.
- . *Apache Nutch*. Accessed October 14, 2013. <http://nutch.apache.org/>.
- . *Apache Oozie*, 09:13:35. <http://oozie.apache.org/>.
- . *Apache Pig*, 20:37:09. <http://pig.apache.org/>.
- . *Apache S4*, n.d. <http://incubator.apache.org/s4/>.
- . *Apache Samza*, 10:42:52. <http://samza.incubator.apache.org/>.
- . *Apache Solr*, 09:15:23. <http://lucene.apache.org/solr/>.
- . *Apache Spark*, n.d. <http://spark-project.org/>.
- . *Apache Sqoop*, n.d. <http://sqoop.apache.org/>.
- . *Apache Tez*, 09:37:54. <http://tez.incubator.apache.org/>.
- . *Apache Thrift*, 11:45:38. <http://thrift.apache.org/>.
- . *Apache ZooKeeper*, 09:14:27. <http://zookeeper.apache.org/>.
- . "SolrEcosystem," 23:05:31. <http://wiki.apache.org/solr/SolrEcosystem>.
- . "The Apache HBase Reference Guide," 21:20:48.  
<http://hbase.apache.org/book/book.html>.
- Apt, Krzysztof R., Evangelos Markakis, and Sunil Simon. *Paradoxes in Social Networks with Multiple Products*. ArXiv e-print, January 31, 2013. <http://arxiv.org/abs/1301.7592>.
- Archive Team. "The WARC Ecosystem." Accessed October 14, 2013.  
[http://www.archiveteam.org/index.php?title=The\\_WARC\\_Ecosystem](http://www.archiveteam.org/index.php?title=The_WARC_Ecosystem).
- Armstrong, Joe. "Solving the Wrong Problem," n.d.  
<http://joearms.github.com/2013/03/28/solving-the-wrong-problem.html>.
- Arthur, David. "Pig and HBase with LucidWorks Big Data," 20:07:13.  
<http://info.lucidworks.com/blog/bid/272738/Pig-and-HBase-with-LucidWorks-Big-Data>.

Atbrox. *Atbr*, 21:56:09. <https://github.com/atbrox/atbr>.

Auradkar, Aditya, Chavdar Botev, Shirshanka Das, Dave De Maagd, Alex Feinberg, Phanindra Ganti, Lei Gao, Bhaskar Ghosh, Kishore Gopalakrishna, and Brendan Harris. "Data Infrastructure at LinkedIn." In *Data Engineering (ICDE), 2012 IEEE 28th International Conference On*, 1370–1381, 2012.  
[http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=6228206](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6228206).

Aurelius. *Faunus*, 14:25:07. <https://github.com/thinkaurelius/faunus>.  
 ———. *Titan*, 14:24:57. <https://github.com/thinkaurelius/titan>.

Automic (UC4). *ONE Automation*, n.d. <http://automic.com/the-automation-engine/>.

Bahmani, Bahman. "Sketching Techniques for Real-time Big Data." 00:19:56.  
<http://strataconf.com/strata2013/public/schedule/detail/27311>.

Ballabio, Luigi. *QuantLib*, 21:03:53. <https://github.com/lballabio/quantlib>.  
 ———. *QuantLib-Python*, 17:44:29. <https://pypi.python.org/pypi/QuantLib-Python>.

Barber, David. *Bayesian Reasoning and Machine Learning*. Cambridge; New York: Cambridge University Press, 2011.

Barlow, Mike. *The Culture of Big Data*. O'Reilly Media, October 8, 2013.  
<http://oreilly.com/data/culture-of-big-data.csp>.

Bartlett, Randy. *A Practitioner's Guide to Business Analytics: Using Data Analysis Tools to Improve Your Organization's Decision Making and Strategy*. New York: McGraw-Hill, 2013.

Bauer, Stefan. *Getting Started With Amazon Redshift*. Packt Publishing, 2013.  
<http://www.myilibrary.com?id=498652>.

Bayer, Judy, and Marie Taillard. "Data Analysis Should Be a Social Event." *Harvard Business Review*, 07:33:18. <http://blogs.hbr.org/2013/04/data-analysis-should-be-a-soci/>.

Beazley, David M. *Python Cookbook*. Beijing: O'Reilly, 2013.  
 ———. *Python Essential Reference*. 4th ed. Developer's Library. Upper Saddle River, NJ: Addison-Wesley, 2009.

Beckett, Dave. *Redland RDF Libraries*, n.d. <http://librdf.org/>.

Bejeck, Bill. *Getting Started with Google Guava Write Better, More Efficient Java, and Have Fun Doing So*. Olton Birmingham [England]: Packt Pub. Ltd., 2013.  
<http://site.ebrary.com/id/10747042>.

Bergstein, Brian. "Review: The Problem with Our Data Obsession." *MIT Technology Review*, 22:49:47.  
<http://www.technologyreview.com/review/511176/the-problem-with-our-data-obsession/>.

Bhattacharya, Dibyendu. "Using Coprocessors to Index Columns in an Elasticsea..." *Technology*, August 2, 2013.  
<http://www.slideshare.net/cloudera/hbasecon-2013-apache-hadoop-and-apache-hbase-for-realtime-video-analytics>.

Biddulph, Matt. *Pig RedisStorer*, 10:31:28. <https://github.com/mattb/pig-redis>.

Binns, Aaron. "Web Analytics with Hadoop and Pig." n.d.  
<http://infolab.stanford.edu/wac/aaronBinnsTalk2012/index.html#slide1>.

Bird, Steven. *Natural Language Processing with Python*. 1st ed. Beijing ; Cambridge [Mass.]: O'Reilly, 2009.  
 ———. *NLTK*, n.d. <http://nltk.org/>.

Bishop, Christopher M. *Pattern Recognition and Machine Learning*. New York, NY: Springer, 2009.

- bitly. *Dablooms*, 14:32:56. <https://github.com/bitly/dablooms>.
- . *Forget-Table*, 14:31:21. <https://github.com/bitly/forgettable>.
- Bladt, Jeff, and Bob Filbin. “A Data Scientist’s Real Job: Storytelling,” 13:00:19. <http://blogs.hbr.org/2013/03/a-data-scientists-real-job-sto/>.
- BNOSAC. “Massive Online Data Stream Mining with R,” 19:49:04. <http://www.bnosc.be/index.php/blog/26-massive-online-data-stream-mining-with-r>.
- Bogaert, Mathias. *Hadoop Ansible Playbook*. Accessed October 17, 2013. <https://github.com/analytically/hadoop-ansible>.
- Borasky, Edward. *Computational-Journalism-Publishers-Workbench*, 09:13:48. <https://github.com/znmeb/Computational-Journalism-Publishers-Workbench>.
- Bosteels, Klaas. *Dumbo*, 10:50:33. <https://github.com/klbostee/dumbo>.
- Bostock, Michael. “Bl.ocks.org - Mbostock,” n.d. <http://bl.ocks.org/mbostock>.
- . *D3 Plugins*, 06:30:57. <https://github.com/d3/d3-plugins>.
- . *D3.js*, 20:18:30. <https://github.com/mbostock/d3>.
- . “How Selections Work,” n.d. <http://bost.ocks.org/mike/selection/>.
- . “Let’s Make a Map,” n.d. <http://bost.ocks.org/mike/map/>.
- . *Topojson*, 10:51:59. <https://github.com/mbostock/topojson>.
- Bostock, Michael, Vadim Ogievetsky, and Jeffrey Heer. “D 3: Data-Driven Documents” (20:20:43). <http://vis.stanford.edu/files/2011-D3-InfoVis.pdf>.
- Botev, Chavdar. “Open Sourcing Databus: LinkedIn’s Low Latency Change Data Capture System.” Accessed October 17, 2013. <http://engineering.linkedin.com/data-replication/open-sourcing-databus-linkedins-low-latency-change-data-capture-system>.
- Botzum, Keys. “Basic Notes on Configuring Eclipse as a Hadoop Development Environment for MapR,” n.d. <http://www.mapr.com/blog/basic-notes-on-configuring-eclipse-as-a-hadoop-development-environment-for-mapr>.
- Bowen, Jonathan. *Getting Started with Talend Open Studio for Data Integration*. Packt Publishing Ltd, 2012.
- Boyd, Danah, and Kate Crawford. *Six Provocations for Big Data*. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, September 21, 2011. <http://papers.ssrn.com/abstract=1926431>.
- Braun, Mikio L. “Big Data Beyond MapReduce: Google’s Big Data Papers.” *Marginally Interesting*. Accessed October 20, 2013. <http://blog.mikiobraun.de/2013/02/big-data-beyond-map-reduce-googles-papers.html>.
- . “More Google Big Data Papers: Megastore and Spanner.” *Marginally Interesting*. Accessed October 20, 2013. <http://blog.mikiobraun.de/2013/03/more-google-papers-megastore-spanner-voted-commits.html>.
- . “The Real-Time Big Data Landscape.” *Marginally Interesting*. Accessed October 20, 2013. <http://blog.mikiobraun.de/2013/06/real-time-big-data-landscape.html>.
- . “Why You Don’t Want Real-time Analytics to Be Exact.” *Marginally Interesting*. Accessed October 20, 2013. <http://blog.mikiobraun.de/2012/08/why-you-dont-want-real-time-analytics-to-be-exact.html>.
- Brewer, Eric. “CAP Twelve Years Later: How the ‘Rules’ Have Changed,” 07:23:39.

- <http://www.infoq.com/articles/cap-twelve-years-later-how-the-rules-have-changed>.
- Britcher, Robert N. *The Limits of Software: People, Projects, and Perspectives*. Reading, Mass.: Addison-Wesley, 1999.
- Brown, Aaron. *Red-blooded Risk: The Secret History of Wall Street*. Hoboken, N.J: Wiley, 2012.
- Bryer, Jason. *Sqlutils*, 19:11:04. <https://github.com/jbryer/sqlutils>.
- Bumgarner, Vincent. *Implementing Splunk: Big Data Reporting and Development for Operational Intelligence*. Birmingham: Packt Pub., 2013.
- Busch, Michael, Krishna Gade, Brian Larson, Patrick Lok, Samuel Luckenbill, and Jimmy Lin. “Earlybird: Real-time Search at Twitter.” In *Data Engineering (ICDE), 2012 IEEE 28th International Conference On*, 1360–1369, 2012.  
[http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=6228205](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6228205).
- Cairo, Alberto. *The Functional Art: An Introduction to Information Graphics and Visualization*, 2013.
- Canny, John, and Huasha Zhao. “Big Data Analytics with Small Footprint: Squaring the Cloud.” In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 95–103, 2013. <http://dl.acm.org/citation.cfm?id=2487677>.
- Capriolo, Edward. *Cassandra High Performance Cookbook*. Birmingham, UK: Packt Pub., 2011. <http://proquest.safaribooksonline.com/?fpi=9781849515122>.
- . *Clojure Data Analysis Cookbook*. [S.l.]: Packt Publishing Limited, 2013.  
<http://lib.myilibrary.com?id=475335>.
- Capriolo, Edward, Dean Wampler, and Jason Rutherglen. *Programming Hive*. Sebastopol, CA: O’Reilly & Associates, 2012.
- Carlsson, Gunnar. “Topology and Data.” *Bulletin of the American Mathematical Society* 46, no. 2 (2009): 255–308.
- Cataldo, Adam. “Quick Deploy: a Distributed Systems Approach to Developer Productivity.” Accessed October 17, 2013.  
<http://engineering.linkedin.com/developer-productivity/quick-deploy-distributed-systems-approach-developer-productivity>.
- Chang, Winston. *R Graphics Cookbook*. Sebastopol, CA: O’Reilly Media, 2012.  
<http://proquest.safaribooksonline.com/?fpi=9781449363086>.
- Chinchilla, Carlos. *Tales*, 22:00:25. <https://github.com/calufa/tales-core>.
- Ching, Avery. “Scaling Apache Giraph to a Trillion Edges,” n.d.  
<https://www.facebook.com/notes/facebook-engineering/scaling-apache-giraph-to-a-trillion-edges/10151617006153920>.
- Chu-Carroll, Mark C. *Good Math: a Geek’s Guide to the Beauty of Numbers, Logic, and Computation*. Dallas, TX: Pragmatic Programmers, 2013.
- Citus Data. *CitusDB*, 21:51:57. <http://www.citusdata.com/downloads>.
- Clarke, Daoud. *Seven Steps to Success: Machine Learning in Practice*, n.d.  
<http://daoudclarke.github.io/guide.pdf>.
- Clayton, Rich. “Why Business Leaders Must Master Data And Analytics.” *Forbes*, 11:19:55.  
<http://www.forbes.com/sites/oracle/2013/04/22/why-business-leaders-must-master-data-and-analytics/>.
- Clegg, Andrew. “Pig-data-mining-talk.” 20:33:09.  
<https://github.com/andrewclegg/pig-data-mining-talk>.
- Cloudera. *Cloudera ML*. Accessed October 20, 2013. <https://github.com/cloudera/ml>.
- . *HTrace*, 12:52:47. <https://github.com/cloudera/htrace>.

- . *Hue*, 16:36:31. <https://github.com/cloudera/hue>.
- . *Impala*, 09:02:47.  
<http://www.cloudera.com/content/cloudera/en/products/cdh/impala.html>.
- . *RecordBreaker*, n.d. <http://cloudera.github.com/RecordBreaker/>.
- Collie, Rob. *DAX Formulas for PowerPivot*. Uniontown, Ohio: Holy Macro! Books, 2013.  
<http://www.books24x7.com/marc.asp?bookid=50324>.
- CommonCrawl. *CommonCrawl Crawler*. Accessed October 14, 2013.  
<https://github.com/commoncrawl/commoncrawl-crawler>.
- Concurrent. *Cascading*, n.d. <http://www.cascading.org/>.
- Condie, Tyson, Neil Conway, Peter Alvaro, Joseph M. Hellerstein, Khaled Elmeleegy, and Russell Sears. “MapReduce Online.” 21–21. NSDI’10. Berkeley, CA, USA: USENIX Association, 2010. <http://dl.acm.org/citation.cfm?id=1855711.1855732>.
- Congiu, Roberto. *Hive-JSON-Serde*, 10:18:48. <https://github.com/rcongiu/Hive-JSON-Serde>.
- Continuuity. *Continuuity Reactor*, 12:30:48. <http://continuuity.com/>.
- . *Weave*, 17:30:30. <https://github.com/continuuity/weave>.
- Continuum Analytics. *Anaconda*. Continuum Analytics, 12:48:47.  
<https://store.continuum.io/cshop/anaconda/>.
- . *Blaze*. Continuum Analytics, Inc., 12:48:25. <https://github.com/ContinuumIO/blaze>.
- . *Bokeh*. Continuum Analytics, Inc., 12:48:35. <https://github.com/ContinuumIO/Bokeh>.
- . *Wakari*, n.d. <https://wakari.io/>.
- Cook, John D. “Recognizing Numbers.” *The Endeavour*, n.d.  
<http://www.johndcook.com/blog/2013/04/30/recognizing-numbers/>.
- Corbett, James C., Jeffrey Dean, Michael Epstein, Andrew Fikes, Christopher Frost, J. J. Furman, Sanjay Ghemawat, Andrey Gubarev, Christopher Heiser, and Peter Hochschild. “Spanner: Google’s Globally-distributed Database.” In *Proceedings of OSDI*. Vol. 1, 2012. <https://www.usenix.org/system/files/conference/osdi12/osdi12-final-16.pdf>.
- Corr, Lawrence, and Jim Stagnitto. *Agile Data Warehouse Design: Collaborative Dimensional Modeling, from Whiteboard to Star Schema*. Leeds: Decisionone Press, 2012.
- Corum, Jonathan. “Storytelling with Data.” n.d. <http://style.org/tapestry/>.
- Craig, Terence, and Mary E Ludloff. *Privacy and Big Data*. Sebastopol, CA: O’Reilly, 2011.
- Crawford, Kate. “The Hidden Biases in Big Data.” *HBR Blog Network - Harvard Business Review*, 15:48:03. <http://blogs.hbr.org/2013/04/the-hidden-biases-in-big-data/>.
- Croll, Alistair. “The Business Singularity: Why Software Means Cycle Time Trumps Scale.” 00:25:54. <http://strataconf.com/strata2013/public/schedule/detail/27993>.
- CRS4. *Pydoop*, 21:26:06. <https://github.com/crs4/pydoop>.
- Damarla, Chanu. “Warren Buffett, The Human Big Data Engine.” *Business 2 Community*, 08:30:54.  
<http://www.business2community.com/big-data/warren-buffett-the-human-big-data-engine-0484510>.
- Daniel Tunkelang. “Find and Be Found: Information Retrieval at LinkedIn.” Technology, July 31, 2013.  
<http://www.slideshare.net/dtunkelang/find-and-be-found-information-retrieval-at-linked-in>.
- Daniels, Doug. “Domesticating Pig with Lipstick.” *Mortar Data Blog*, 20:20:17.  
<http://blog.mortardata.com/post/60765120319/domesticating-pig-with-lipstick>.
- Das, Tathagata. “Deep Dive with Spark Streaming.” June 23, 2013.

- <http://www.slideshare.net/spark-project/deep-divewithsparkstreaming-tathagatadasspark-meetup20130617>.
- Datameer. *Datameer*, n.d. <http://www.datameer.com/index.html>.
- Datasalt. *Pangool*, 21:16:14. <http://pangool.net/>.
- Davenport, Thomas H. *Enterprise Analytics: Optimize Performance, Process and Decisions through Big Data*. Upper Saddle River, N.J.: Financial Times/Prentice Hall, 2013.
- Davenport, Thomas H, and Jinho Kim. *Keeping up with the Quants: Your Guide to Understanding and Using Analytics*, 2013.
- Davenport, Tom. "How P&G Presents Data to Decision-Makers," n.d. [http://blogs.hbr.org/cs/2013/04/how\\_p\\_and\\_g\\_presents\\_data.html](http://blogs.hbr.org/cs/2013/04/how_p_and_g_presents_data.html).
- Davidson-Pilon, Cameron. "Probabilistic Programming and Bayesian Methods for Hackers." *GitHub*, 07:20:43. <https://github.com/CamDavidsonPilon/Probabilistic-Programming-and-Bayesian-Methods-for-Hackers>.
- Davis, Kord, and Doug Patterson. *Ethics of Big Data*. Farnham: O'Reilly, 2012.
- Day, Allen. "Data Platform Design Patterns." Education, September 17, 2013. <http://www.slideshare.net/allenday54/20130905-singapore-big-datavg>.
- . "NoSQL and SQL Work Side-by-Side to Tackle Real-time Big Data Needs." September 17, 2013. <http://www.slideshare.net/allenday54/20130617-new-york-open-analytics-summit-no-sql-and-sql>.
- De Cadenza, Michele. "Online Available Data Services: a Primer." Education, July 3, 2013. <http://www.slideshare.net/indecadenza/daa-s-xchange2013galassopiuntiv04>.
- De Goes, John. "Analytics Maturity Model." Technology, April 9, 2013. <http://www.slideshare.net/jdegoes/analytics-maturity-model>.
- Deepspeed media. *Internet Archive*, 2013. <http://vimeo.com/59207751>.
- Deming, W. Edwards. *Out of the Crisis*. Cambridge, Mass: Massachusetts Institute of Technology, Center for Advanced Engineering Study, 1986.
- Denis, Frank. "Why We Love Apache Pig." *Umbrella Security Labs*, 20:10:20. <http://labs.umbrella.com/2013/04/08/pig-jruby/>.
- Devlin, Catherine. *Ipython-sql*, 14:29:32. <https://pypi.python.org/pypi/ipython-sql>.
- DigitalPebble. *Behemoth*, 21:44:24. <https://github.com/DigitalPebble/behemoth>.
- Dimiduk, Nick. "HBase Client APIs (for Webapps?)." March 28, 2013. <http://www.slideshare.net/xfyr/hbase-client-apis-for-webapps>.
- . *HBase in Action*. Shelter Island, NY: Manning, 2013.
- Disco Project. *Disco*, 18:30:30. <https://github.com/discoproject/disco>.
- Downey, Allen. *Think Python*. Sebastopol, CA: O'Reilly, 2012. <http://www.contentreserve.com/TitleInfo.asp?ID={8BDDEF5E-F64A-4212-A868-9B60E742EC40}&Format=50>.
- . *Think Stats*. 1st ed. Probability and Statistics for Programmers. Sebastopol, CA: O'Reilly Media, 2011.
- Downey, Allen B. *Think Complexity*. Beijing [etc.]: O'Reilly, 2012.
- Downey, Allen B. *Think Bayes*. Oreilly & Associates Inc, 2013.
- Duarte, Nancy. "When Presenting Your Data, Get to the Point Fast." *HBR Blog Network - Harvard Business Review*, 11:18:43. <http://blogs.hbr.org/2013/03/when-presenting-your-data-get/>.

- DuCharme, Bob. *Learning SPARQL: Querying and Updating with SPARQL 1.1*. 1st ed. Sebastopol, CA: O'Reilly Media, 2011.
- Dunn, Roland, Ger Hobbelt, Andrew Thornton, Chris Viau, and Troy Mott. *Developing a D3.js Edge*. Bleeding Edge Press, 2013.
- Dunning, Ted. "Introduction to Mahout and How To Build a Recommender." Technology, August 24, 2013.  
<http://www.slideshare.net/tdunning/dfw-big-data-talk-on-mahout-recommenders>.
- . "Multi-Modal Recommendation." Technology, June 5, 2013.  
<http://www.slideshare.net/tdunning/buzz-wordsdunningmultimodalrecommendation>.
- . "Using Mahout and a Search Engine for Recommendation." Technology, October 5, 2013.  
<http://www.slideshare.net/tdunning/using-mahout-and-a-search-engine-for-recommendation>.
- eBay Software Foundation. *Ql.io*, 22:08:08. <https://github.com/ql-io/ql.io>.
- Eckerson, Wayne. "Analytical Modeling Is Both Science and Art," 12:49:54.  
<http://searchbusinessanalytics.techtarget.com/opinion/Analytical-modeling-is-both-science-and-art>.
- . *Secrets of Analytical Leaders: Insights from Information Insiders*. 1st ed. Westfield, NJ: Technics Publications, LLC, 2012.
- . *Visual Discovery Tools: Market Segmentation and Product Positioning*, 00:50:08.  
[http://www.idgconnect.com/view\\_abstract/14386/visual-discovery-tools-market-segmentation-product-positioning](http://www.idgconnect.com/view_abstract/14386/visual-discovery-tools-market-segmentation-product-positioning).
- Elasticsearch. *ElasticSearch*, 21:48:26. <https://github.com/elasticsearch/elasticsearch>.
- . *Elasticsearch Hadoop*, 21:46:29. <https://github.com/elasticsearch/elasticsearch-hadoop>.
- Elmahrek, Abraham. "How-to: Import a Pre-existing Oozie Workflow into Hue," n.d.  
<http://blog.cloudera.com/blog/2013/03/how-to-import-a-pre-existing-oozie-workflow-into-hue/>.
- EMRG. *NodeBox*, n.d. <http://nodebox.net/>.
- Enthought. *EPD*, n.d. <http://www.enthought.com/>.
- EPFL. *Scala*, 12:42:47. <http://www.scala-lang.org/>.
- Epstein, Charles, Gunnar Carlsson, and Herbert Edelsbrunner. "Topological Data Analysis." *Inverse Problems* 27, no. 12 (December 1, 2011): 120201.  
doi:10.1088/0266-5611/27/12/120201.
- Erickson, Justin. "Impala: A Modern SQL Engine for Hadoop." 00:17:11.  
<http://strataconf.com/strata2013/public/schedule/detail/27991>.
- EsperTech. *Esper*, n.d. <http://esper.codehaus.org/>.
- Esri. *GIS Tools for Hadoop*, 10:24:08. <https://github.com/Esri/gis-tools-for-hadoop>.
- . *Spatial Framework for Hadoop*, 10:23:52.  
<https://github.com/Esri/spatial-framework-for-hadoop>.
- European Journalism Centre, and Open Knowledge Foundation. *The Data Journalism Handbook*. 1st ed. Sebastopol, CA: O'Reilly Media, 2012.
- Facebook. *Scribe*, 21:55:19. <https://github.com/facebook/scribe>.
- Farina, Daniel. *WAL-E*, n.d. <https://github.com/wal-e/wal-e#wal-e-disaster-recovery>.
- Farmer, F. Randall, and Bryce Glass. *Building Web Reputation Systems*. Sebastopol, CA: O'Reilly, 2010.
- Faroult, Stéphane. *The Art of SQL*. 1st ed. Sebastopol, CA: O'Reilly, 2006.



- Ferrera, Pedro, Ivan de Prado, Eric Palacios, Jose Luis Fernandez-Marquez, and Giovanna Di Marzo Serugendo. "Tuple MapReduce: Beyond Classic MapReduce." In *Data Mining (ICDM), 2012 IEEE 12th International Conference On*, 260–269, 2012.  
[http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=6413897](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6413897).
- Few, Stephen. *Information Dashboard Design: Displaying Data for At-a-glance Monitoring*. Burlingame, CA: Analytics Press, 2013.
- . *Information Dashboard Design: The Effective Visual Communication of Data*. 1st ed. Beijing ; Cambride [MA]: O'Reilly, 2006.
- . *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. Oakland, Calif: Analytics Press, 2009.
- Flach, Peter. *Machine Learning*. Cambridge: CAMBRIDGE UNIVERSITY PRESS, 2012.
- Flach, Peter A. *Machine Learning the Art and Science of Algorithms That Make Sense of Data*. Cambridge: Cambridge University Press, 2012.  
<http://dx.doi.org/10.1017/CBO9780511973000>.
- Fonnesbeck, Chris, Anand Patil, and John Salvatier. *PyMC*, 20:59:34.  
<https://github.com/pymc-devs/pymc>.
- Ford, Daniel, Fran\ccois Labelle, Florentina I. Popovici, Murray Stokely, Van-Anh Truong, Luiz Barroso, Carrie Grimes, and Sean Quinlan. "Availability in Globally Distributed Storage Systems." In *OSDI*, 61–74, 2010.  
[http://static.usenix.org/events/osdi10/tech/full\\_papers/Ford.pdf](http://static.usenix.org/events/osdi10/tech/full_papers/Ford.pdf).
- Franks, Bill. *Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics*, 2012.
- Fredheim, Rolf. "Quantifying Memory: Experiments in Python and D3 from R: GDELT Made Easy." *Quantifying Memory*, 22:09:31.  
<http://quantifyingmemory.blogspot.com/2013/04/experiments-in-python-and-d3-from-r.html>.
- "Free-programming-books." *GitHub*. Accessed October 17, 2013.  
<https://github.com/vhf/free-programming-books>.
- Fung, Kaiser. *Numbersense: How to Use Big Data to Your Advantage*, 2013.
- Furuhashi, Sadayuki. *MessagePack*, 19:52:05. <https://github.com/msgpack/msgpack>.
- . *MessagePack-Hadoop*, 19:52:42. <https://github.com/msgpack/msgpack-hadoop>.
- Galloway, Alexander R. *The Interface Effect*. Cambridge, UK ; Malden, MA: Polity, 2012.
- Ganglia Development Team. *Ganglia Monitoring System*, 14:17:19. <http://ganglia.info/>.
- Ganjisaffar, Yasser. *Crawler4j*. Accessed October 14, 2013.  
<https://code.google.com/p/crawler4j/>.
- García, Miguel, and Barry Harmsen. *QlikView 11 for Developers: Develop Business Intelligence Applications with QlikView 11*. Birmingham: Packt Pub., 2012.
- Garnaat, Mitch. *Python and AWS Cookbook*. Sebastopol, CA: O'Reilly, 2012.  
<http://proquest.safaribooksonline.com/9781449308100>.
- Gates, Alan. "How to Make Apache Hive 100X Faster? Apache Hive Stinger." *Hortonworks*, 21:00:13. <http://hortonworks.com/blog/100x-faster-hive/>.
- . "Pig and Hive at Yahoo!" *Yahoo*, 20:11:58.  
<http://developer.yahoo.com/blogs/hadoop/pig-hive-yahoo-464.html>.
- . *Programming Pig*. Sebastopol, CA: O'Reilly, 2011.  
<http://proquest.safaribooksonline.com/?fpi=9781449317881>.
- Gates, Alan F., Olga Natkovich, Shubham Chopra, Pradeep Kamath, Shravan M.

- Narayanamurthy, Christopher Olston, Benjamin Reed, Santhosh Srinivasan, and Utkarsh Srivastava. "Building a High-level Dataflow System on Top of Map-Reduce: The Pig Experience." *Proc. VLDB Endow.* 2, no. 2 (August 2009): 1414–1425.
- George, Lars. *HBase the Definitive Guide*. Sebastopol, CA: O'Reilly, 2011.  
<http://proquest.safaribooksonline.com/9781449314682>.
- Gitelman, Lisa. "*Raw Data*" Is an Oxymoron, 2013.  
<http://books.google.com/books?isbn=0262518287>.
- Goodhope, Ken, Joel Koshy, Jay Kreps, Neha Narkhede, Richard Park, Jun Rao, and Victor Yang Ye. "Building LinkedIn's Real-time Activity Data Pipeline." *IEEE Data Eng. Bull.* 35, no. 2 (2012): 33–45.
- Goodman, Sasha C. *RSeek.org*, n.d. <http://rseek.org/>.
- Google. *Google BigQuery*, n.d. <https://developers.google.com/bigquery/docs/overview>.  
 ———. *Google Fusion Tables*, 12:40:34. <https://developers.google.com/fusiontables/>.  
 ———. *Protocol Buffers*, 09:45:37. <https://code.google.com/p/protobuf/>.
- Gopalakrishna, Kishore. "Untangling Cluster Management with Helix." January 29, 2013.  
<http://www.slideshare.net/amywtang/helix-socc-v10final-16247096>.
- Goyal, Sonal. *Crux*, 00:51:45. <https://github.com/sonalgoyal/crux>.  
 ———. *HIHO*, 12:12:45. <https://github.com/sonalgoyal/hiho>.
- Graham, Bill. "Hadoop and Vertica: The Data Analytics Platform at Twitter." *Business & Mgmt*, June 21, 2012.  
<http://www.slideshare.net/billonahill/hadoop-summit-2012-hadoop-and-vertica-the-data-analytics-platform-at-twitter>.
- Gravity. *HPaste*, 21:52:42. <https://github.com/GravityLabs/HPaste>.
- Gregg, Brendan. "Linux Performance Analysis and Tools." March 1, 2013.  
<http://www.slideshare.net/brendangregg/linux-performance-analysis-and-tools>.
- Gregg, Forest, and Derek Eder. *Dedupe*, 21:21:44. <https://github.com/open-city/dedupe>.
- Groskopf, Christopher. *Csvkit*, 06:35:31. <https://github.com/onyxfish/csvkit>.
- Guha, Saptarshi. *RHIPE*, 09:23:09. <http://www.datadr.org/>.
- Guha, Saptarshi, Ryan Hafen, Jeremiah Rounds, Jin Xia, Jianfu Li, Bowei Xi, and William S. Cleveland. "Large Complex Data: Divide and Recombine (D&R) with RHIPE." *Stat* 1, no. 1 (2012): 53–67.
- Guha, Saptarshi, Paul Kidwell, Ryan Hafen, and William S. Cleveland. "Visualization Databases for the Analysis of Large Complex Datasets." 193–200, 2009.  
[http://machinelearning.wustl.edu/mlpapers/paper\\_files/AISTATS09\\_GuhaKHC.pdf](http://machinelearning.wustl.edu/mlpapers/paper_files/AISTATS09_GuhaKHC.pdf).
- Guo, Shumin. *Hadoop Operations and Cluster Management Cookbook*. Birmingham, UK: Packt Publishing Ltd., 2013. <http://site.ebrary.com/id/10742618>.
- Gupta, Pankaj, Ashish Goel, Jimmy Lin, Aneesh Sharma, Dong Wang, and Reza Zadeh. "Wtf: The Who to Follow Service at Twitter." In *Proceedings of the 22nd International Conference on World Wide Web*, 505–514, 2013.  
<http://dl.acm.org/citation.cfm?id=2488433>.
- Hafen, Ryan. *Datadr*, 09:26:50. <https://github.com/hafen/datadr>.  
 ———. *Trelliscope*, 09:24:30. <https://github.com/hafen/trelliscope>.
- Hall, Alexander, Olaf Bachmann, Robert Büsow, Silviu Guanceanu, and Marc Nunkesser. "Processing a Trillion Cells Per Mouse Click." *Proceedings of the VLDB Endowment* 5, no. 11 (2012): 1436–1446.
- Hamstra, Mark, and Matei Zaharia. *Learning Spark: Lightning-fast Big Data Analytics*, 2013.

- Hanborq. *PigML*, 11:09:46. <https://github.com/hanborq/pigml>.
- . *RockStor*, 11:10:01. <https://github.com/hanborq/rockstor>.
- Hargrave, Vic. “Securing Hadoop with OSSEC.” Accessed October 17, 2013. <http://vichargrave.com/securing-hadoop-with-ossec/>.
- Harrington, Peter. *Machine Learning in Action*. Shelter Island, N.Y.: Manning Publications Co., 2012.
- Hastie, Trevor, Robert Tibshirani, and J. H Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer, 2009.
- Hausenblas, Michael. “Apache Drill: Interactive Analytics for Large--scale Datasets.” Accessed October 17, 2013. <https://speakerdeck.com/mhausenblas/hug-stockholm-apache-drill>.
- Hayes, Matthew. “A Brief Tour of DataFu.” Design, June 25, 2013. <http://www.slideshare.net/matthewterencehayes/datafu>.
- . “Building Data Products at LinkedIn with DataFu.” Technology, August 26, 2013. <http://www.slideshare.net/matthewterencehayes/building-data-products-at-linkedin-with-datafu>.
- . “Hourglass: a Library for Incremental Processing on Hadoop.” Education, October 9, 2013. <http://www.slideshare.net/matthewterencehayes/hourglass-a-library-for-incremental-processing-on-hadoop>.
- Hayes, Matthew, and Sam Shah. “Hourglass: a Library for Incremental Processing on Hadoop.” Santa Clara, CA, USA, n.d.
- Hearst, Marti. “Analyzing Big Data with Twitter | A Special UC Berkeley iSchool Course.” Accessed October 15, 2013. <http://blogs.ischool.berkeley.edu/i290-abdt-s12/2012/12/13/uc-berkeley-course-lectures-analyzing-big-data-with-twitter/>.
- Heer, Jeffrey, Michael Bostock, and Vadim Ogievetsky. “A Tour Through the Visualization Zoo,” 12:34:43. <http://hci.stanford.edu/jheer/files/zoo/>.
- Helmling, Gary. “A Bird’s-Eye View of Pig and Scalding with hRaven.” Business & Mgmt, June 28, 2013. <http://www.slideshare.net/ghelmling/helmling-june27-1100amhall1?ref=http://eventifier.co/event/hadoops summit2013/slides>.
- Hemann, Chuck, and Ken Burbary. *Digital Marketing Analytics Making Sense of Consumer Data in a Digital World*. Indianapolis, Ind.: Que, 2013. <http://proquest.safaribooksonline.com/?fpi=9780133150933>.
- Herodotou, Herodotos, Harold Lim, Gang Luo, Nedyalko Borisov, Liang Dong, Fatma Bilgen Cetin, and Shivnath Babu. “Starfish: A Self-tuning System for Big Data Analytics.” 11:261–272, 2011. [http://x86.cs.duke.edu/~gang/documents/CIDR11\\_Paper36.pdf](http://x86.cs.duke.edu/~gang/documents/CIDR11_Paper36.pdf).
- Herodotou, Herodotos, Harold Lim, Gang Luo, Nedyalko Borisov, Liang Dong, Fatma Bilgen Cetin, Shivnath Babu, Jie Li, Kunal Agarwal, and Fei Dong. *Starfish*, 13:22:13. <http://www.cs.duke.edu/starfish/>.
- Heule, Stefan, Marc Nunkesser, and Alexander Hall. “HyperLogLog in Practice: Algorithmic Engineering of a State of the Art Cardinality Estimation Algorithm.” 683–692, 2013. <http://dl.acm.org/citation.cfm?id=2452456>.
- Hewlett-Packard. *Vertica*, 09:00:24. <http://www.vertica.com/>.
- Hey, Tony, and Kristin Tolle. *The Fourth Paradigm Data-intensive Scientific Discovery*. Redmond, Wash.: Microsoft Research, 2009.

- Hindman, Benjamin, Andy Konwinski, Matei Zaharia, Ali Ghodsi, Anthony D. Joseph, Randy Katz, Scott Shenker, and Ion Stoica. "Mesos: A Platform for Fine-grained Resource Sharing in the Data Center." In *Proceedings of the 8th USENIX Conference on Networked Systems Design and Implementation*, 22–22, 2011.  
[http://static.usenix.org/events/nsdi11/tech/full\\_papers/Hindman\\_new.pdf](http://static.usenix.org/events/nsdi11/tech/full_papers/Hindman_new.pdf).
- Hintjens, Pieter. *ZeroMQ*. Beijing: O'Reilly, 2013.
- Hirst, Tony. "Mapping Corporate Networks With OpenCorporates." Technology, April 23, 2013.  
<http://www.slideshare.net/psychemedia/mapping-corporate-networks-with-opencorporates?ref=http://blog.ouseful.info/2013/04/23/mapping-corporate-networks-with-opencorporates/>.
- Hoffman, Steve. *Apache Flume: Distributed Log Collection for Hadoop*. Packt Publishing, 2013.  
<http://lib.myilibrary.com?id=504765>.
- Højsgaard, Søren, David Edwards, and Steffen L Lauritzen. *Graphical Models with R*. New York: Springer, 2012. <http://dx.doi.org/10.1007/978-1-4614-2299-0>.
- Holmes, Alex. *Hadoop in Practice*. Shelter Island, NY: Manning, 2012.  
 ———. *Htuple*. Accessed October 17, 2013. <https://github.com/alexholmes/htuple>.
- Hornick, Mark, and Tom Plunkett. *Using R to Unlock the Value of Big Data: Big Data Analytics with Oracle R Enterprise and Oracle R Connector for Hadoop*. New York: McGraw-Hill, 2013.
- Hortonworks. *Apache Hadoop Patterns of Use: Refine, Enrich and Explore*, 00:15:55.  
<http://hortonworks.com/blog/apache-hadoop-patterns-of-use-refine-enrich-and-explore/>.
- . *Gohadoop*. Accessed October 17, 2013. <https://github.com/hortonworks/gohadoop>.
- . "Hadoop Icons," n.d.  
[http://hortonworks.com/blog/a-set-of-hadoop-related-icons/?mkt\\_tok=3RkMMJWWfF9wsRoiua7OZKXonjHpfsX54%2BgvXKG%2FIMI%2F0ER3fOvrPUfGjI4CTsdjI%2BSLDwEYGJlv6SgFT7TMMbFh1rgNUxc%3D](http://hortonworks.com/blog/a-set-of-hadoop-related-icons/?mkt_tok=3RkMMJWWfF9wsRoiua7OZKXonjHpfsX54%2BgvXKG%2FIMI%2F0ER3fOvrPUfGjI4CTsdjI%2BSLDwEYGJlv6SgFT7TMMbFh1rgNUxc%3D).
- . "Hadoop Patterns of Use." Business & Mgmt, April 10, 2013.  
<http://www.slideshare.net/hortonworks/hadoop-patterns-of-use>.
- . "Hive Functions Cheat Sheet." Technology, September 16, 2013.  
<http://www.slideshare.net/hortonworks/hwxqubolehiveudfguide10>.
- . *Hoya*, 21:24:10. <https://github.com/hortonworks/hoya>.
- . *The Business Value of Hadoop as Seen through the Big Data*, 00:25:06.  
<http://hortonworks.com/blog/the-business-value-of-hadoop-as-seen-through-the-big-data/>.
- Howard, Jeremy, Margit Zwemer, and Mike Loukides. *Designing Great Data Products*. O'Reilly Media, Inc., 2012.
- Hsieh, Jonathan, and Himanshu Vashishtha. "Using HBase Effectively - What You Need to Know as an Application Developer." 00:22:50.  
<http://strataconf.com/strata2013/public/schedule/detail/27109>.
- Hubbard, Douglas W. *How to Measure Anything Finding the Value of "Intangibles" in Business*. Hoboken, N.J.: Wiley, 2010. <http://site.ebrary.com/id/10381009>.
- Huber, Peter J. *Data Analysis: What Can Be Learned From the Past 50 Years*. John Wiley & Sons, 2012.
- Huff, Darrell. *How to Lie with Statistics*. New York: Norton, 1993.
- Hultberg, Theo. *Piglet*, 18:20:11. <https://github.com/iconara/piglet>.
- . *Rubydoop*, 11:03:30. <https://github.com/iconara/rubydoop>.

Huß, Roland. *Jolokia*, 12:58:01. <https://github.com/rhuss/jolokia>.

IBM. *Many Eyes*, n.d. <http://www-958.ibm.com/software/data/cognos/manyeyes/>.

Idris, Ivan. *NumPy Cookbook*. Birmingham: Packt Pub., 2012.  
<http://proquest.safaribooksonline.com/?fpi=9781849518925>.

iMatix. *ZeroMQ*, 09:05:54. <http://zeromq.org/>.

Infochimps. *How to Do a Big Data Project*, n.d.  
<http://bigdata.infochimps.com/Portals/174427/docs/bigdataprojecthowto.pdf>.

———. *Piggy*, 20:29:08. <https://github.com/infochimps-labs/piggy>.

———. *Wonderdog*, 10:30:31. <https://github.com/infochimps-labs/wonderdog>.

———. *Wukong*, 12:12:51. <https://github.com/infochimps/wukong>.

Ingersoll, Grant S. “Large-Scale Search Discovery Analytics with Hadoop, Mahout, Solr.”  
 Technology, June 20, 2012.  
[http://www.slideshare.net/Hadoop\\_Summit/large-scale-search-discover-analytics-mahoot-solr](http://www.slideshare.net/Hadoop_Summit/large-scale-search-discover-analytics-mahoot-solr).

———. *Taming Text: How to Find, Organise, and Manipulate It*. Shelter Island, NY : London: Manning ; Pearson Education [distributor], 2013.

Intel OTC. *GraphBuilder*. Accessed October 20, 2013. <https://github.com/01org/graphbuilder>.

Internet Archive. *Heritrix*. Accessed October 12, 2013.  
<https://github.com/internetarchive/heritrix3>.

———. *Waimea*, 20:32:23. <https://github.com/internetarchive/waimea>.

IPython Development Team. *IPython*, 14:28:51. <https://github.com/ipython/ipython>.

Isson, Jean Paul, and Jesse Harriott. *Win with Advanced Business Analytics Creating Business Value from Your Data*. Hoboken, N.J.: John Wiley & Sons, 2013.  
<http://proxy.uqtr.ca/login.cgi?action=login&u=uqtr&db=books24x7&ezproxy=1&ezurl=http://library.books24x7.com/library.asp?%5EB&bookid=46732>.

Jacomy, Alexis. *Sigma.js*, n.d. <http://sigmajavascript.org/>.

Janert, Philipp K. *Data Analysis with Open Source Tools*. Sebastopol, CA: O’Reilly, 2011.

Jannach, Dietmar. *Recommender Systems: An Introduction*. New York: Cambridge University Press, 2011.

Jiang, Y. *HBase Administration Cookbook*. Birmingham: Packt Publishing, Limited, 2012.  
<https://ezproxy.siasat.sk.ca:443/login?url=http://proquest.safaribooksonline.com/9781849517140>.

Johnson, Barry. *Algorithmic Trading & DMA: An Introduction to Direct Access Trading Strategies*. London: 4Myeloma Press, 2010.

Johnson, Grier. “Autometrics: Self-service Metrics Collection.” Accessed October 17, 2013.  
<http://engineering.linkedin.com/52/autometrics-self-service-metrics-collection>.

Johnson, Ian. *Tributary*, 22:04:10. <https://github.com/enjalot/tributary>.

Jonas, Jeff. “Streaming Analytics Vs. Perpetual Analytics (Advantages of Windowless Thinking),” 19:24:09.  
[http://jeffjonas.typepad.com/jeff\\_jonas/2007/04/streaming\\_analy.html](http://jeffjonas.typepad.com/jeff_jonas/2007/04/streaming_analy.html).

Journey, Russell. “HOWTO Use Hive to SQLize Your Own Tweets - Part Two: Loading Hive, SQL Queries.” *Hortonworks*, 20:55:52.  
<http://hortonworks.com/blog/howto-use-hive-to-sqlize-your-own-tweets-part-two-loading-hive-sql-queries/>.

———. “LinkedIn, Apache Pig, and Open Source,” 20:23:28.  
<http://blog.linkedin.com/2010/07/01/linkedin-apache-pig/>.

- . “Pig, ToJson, and Redis to Publish Data with Flask.” *Hortonworks*, 20:22:50.  
<http://hortonworks.com/blog/pig-tojson-and-redis-to-publish-data-with-flask/>.
- JSON-LD. *JSON-LD*, 19:30:23. <https://github.com/json-ld/json-ld.org>.
- Junqueira, Flavio, and Benjamin Reed. *Zookeeper: Distributed Process Coordination*. O’Reilly Media, Incorporated, 2013.
- Jurka, Tim. *RTextTools*, 12:42:00. <https://github.com/timjurka/RTextTools>.
- Jurney, Russell. *Agile Data Science: Building Data Analytics Applications with Hadoop*. [S.l.]: O’Reilly Media, 2012.
- Kabacoff, Robert. *R in Action: Data Analysis and Graphics with R*. Shelter Island, NY: Manning, 2011.
- Kamat, Govind, and Sumeet Singh. “Compression Options in Hadoop - A Tale of Tradeoffs.” July 10, 2013.  
[http://www.slideshare.net/Hadoop\\_Summit/kamat-singh-june27425pmroom210cv2](http://www.slideshare.net/Hadoop_Summit/kamat-singh-june27425pmroom210cv2).
- Kane, Andrew. *Chartkick*, 08:32:42. <https://github.com/ankane/chartkick>.
- Karambelkar, H. *Scaling Big Data with Hadoop and Solr*. [S.l.]: Packt Publishing Limited, 2013.
- Karmasphere. *Karmasphere*, 09:52:22. <http://www.karmasphere.com/>.
- Karnezos, Timon. “AWS Redshift: How Amazon Changed The Game,” n.d.  
<http://blog.aggregateknowledge.com/2013/05/16/aws-redshift-how-amazon-changed-the-game/>.
- Katz, Joel. *Designing Information: Human Factors and Common Sense in Information Design*. Hoboken, N.J: Wiley, 2012.
- Kelly, Jeff. “Big Data Vendor Revenue and Market Forecast 2012-2017,” 23:56:31.  
[http://wikibon.org/w/index.php?title=Big\\_Data\\_Vendor\\_Revenue\\_and\\_Market\\_Forecast\\_2012-2017&printable=yes](http://wikibon.org/w/index.php?title=Big_Data_Vendor_Revenue_and_Market_Forecast_2012-2017&printable=yes).
- Kelly, Mat. *WARCreate*. Accessed October 14, 2013. <https://github.com/machawk1/warcreate>.
- . “WARCreate - Create Wayback-Consumable WARC Files from Any Webpage.” Technology, July 27, 2012.  
<http://www.slideshare.net/matkelly01/warcreate-create-waybackconsumable-warcs-files-from-any-webpage>.
- . *Web Archiving Integration Layer (WAIL)*. Accessed October 14, 2013.  
<https://github.com/machawk1/wail>.
- Kent, William. *Data and Reality*. Bloomington, Ind.: 1st Books Library, 2000.
- Khanwalkar, Manoj, and Govind Asawa. “Evolving a First-Generation Apache HBase Deployment...” Technology, August 2, 2013.  
<http://www.slideshare.net/cloudera/hbasecon-2013-evolving-a-firstgeneration-apache-hbase-deployment-to-second-generation-and-beyond>.
- Kiji Project. *Kiji*, 21:25:56. <http://www.kiji.org/>.
- Kim, Henry M., Iryna Gel, and Ho-Nam Cheung. *Data Analytics Using Ontologies of Management Theories: Towards Implementing “From Theory to Practice.”* SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, August 29, 2012.  
<http://papers.ssrn.com/abstract=2138121>.
- Kimball, Ralph, and Margy Ross. *The Data Warehouse Toolkit The Definitive Guide to Dimensional Modeling*. Hoboken: Wiley, 2013.  
<http://public.eblib.com/EBLPublic/PublicView.do?ptiID=1313513>.
- Kirk, Andy. *Data Visualization a Successful Design Process*. Birmingham, UK: Packt Pub, 2013.

- . “Essential Collection of Visualisation Resources,” 11:26:00.  
<http://www.visualisingdata.com/index.php/resources/>.
- . “The 8 Hats of Data Visualisation.” Technology, June 23, 2012.  
<http://www.slideshare.net/visualisingdata/the-8-hats-of-data-visualisation>.
- . “Understanding Learning in Order to Implement Efficient Visualisation Methods.” Technology, September 21, 2012.  
<http://www.slideshare.net/visualisingdata/andy-kirk-talk-at-big-data-world-europe-september-2012>.
- . “Visualisation Workflow: Finding Stories and Telling Stories.” January 10, 2013.  
<http://www.slideshare.net/visualisingdata/andy-kirks-facebook-talk>.
- Klr. “D3 Lifeline from Vega and Clickme.” *Timely Portfolio*, April 4, 2013.  
<http://timelyportfolio.blogspot.com/2013/04/d3-lifeline-from-vega-and-clickme.html>.
- KNIME.com AG. *KNIME*, 12:25:36. <http://www.knime.org/>.
- Kohei, KaiGai. *Pg\_strom*, 16:29:15. [https://github.com/kaigai/pg\\_strom](https://github.com/kaigai/pg_strom).
- Koomey, Jon. *Turning Numbers into Knowledge: Mastering the Art of Problem Solving*. Oakland, CA: Analytics Press, 2008.
- Kosara, Robert, and Jock Mackinlay. “Storytelling: The Next Step for Visualization.” *Computer* 46, no. 5 (2013): 44–50.
- Koshy, Joel. “Building a Real-Time Data Pipeline: Apache Kafka at LinkedIn.” Technology, July 2, 2013.  
<http://www.slideshare.net/amywtang/building-a-realtime-data-pipeline-apache-kafka-at-linkedin>.
- Kozea. *Multicorn*. Accessed October 20, 2013. <https://github.com/Kozea/Multicorn>.
- Kraska, Tim, Ameet Talwalkar, John C. Duchi, Rean Griffith, Michael J. Franklin, and Michael I. Jordan. “MLbase: A Distributed Machine-learning System.” In *CIDR*, 2013.  
<http://www.cs.berkeley.edu/~ameet/mlbase.pdf>.
- Kreps, Jay. “Getting Real About Distributed System Reliability.” *Jay Kreps*. Accessed October 17, 2013.  
<http://blog.empathybox.com/post/19574936361/getting-real-about-distributed-system-reliability>.
- Krishnan, Sriram. “Genie - Hadoop Platform as a Service at Netflix.” Technology, June 25, 2013. <http://www.slideshare.net/krishflix/genie-hadoop-platform-as-a-service-at-netflix>.
- Kuć, Rafał, and Marek Rogoziński. *Elasticsearch Server: Create a Fast, Scalable, and Flexible Search Solution with the Emerging Open Source Search Server, Elasticsearch*. Birmingham [u.a.]: Packt Publ., 2013.
- Kuhn, Max, and Kjell Johnson. *Applied Predictive Modeling*. New York, NY: Springer, 2013.
- Kumar, Hari. “HBase: Performance Tuners,” 21:21:40.  
<http://labs.ericsson.com/blog/hbase-performance-tuners>.
- Kyrola, Aapo, Guy Blelloch, and Carlos Guestrin. “GraphChi: Large-scale Graph Computation on Just a PC.” 31–46, 2012.  
<https://www.usenix.org/system/files/conference/osdi12/osdi12-final-126.pdf>.
- Kyrola, Aapo, and GraphLab. *GraphChi*, n.d. <http://graphchi.org>.
- Ladley, John. *Data Governance How to Design, Deploy, and Sustain an Effective Data Governance Program*. Waltham, Mass.: Morgan Kaufmann, 2012.
- Lam, Chuck. *Hadoop in Action*. Greenwich, Conn.: Manning Publications, 2011.
- Lamb, Andrew, Matt Fuller, Ramakrishna Varadarajan, Nga Tran, Ben Vandiver, Lyric Doshi,

- and Chuck Bear. "The Vertica Analytic Database: C-store 7 Years Later." *Proceedings of the VLDB Endowment* 5, no. 12 (2012): 1790–1801.
- Lankhorst, Marc. *Enterprise Architecture at Work Modelling, Communication and Analysis*. Berlin, Heidelberg: Springer Berlin Heidelberg : Imprint: Springer, 2013.
- Lans, Rick F. van der. *Data Virtualization for Business Intelligence Systems: Revolutionizing Data Integration for Data Warehouses*. Amsterdam: Elsevier, Morgan Kaufmann, 2012.
- Le Dem, Julien. "Dremel Made Simple with Parquet." *Twitter Blogs*. Accessed October 17, 2013. <https://blog.twitter.com/2013/dremel-made-simple-with-parquet>.
- Le Dem, Julien, and Nong Li. "Parquet: Columnar Storage for the People." *Technology*, June 27, 2013. <http://www.slideshare.net/julienledem/parquet-hadoop-summit-2013>.
- Ledolter, Johannes. *Data Mining and Business Analytics with R*. Hoboken, N.J.: Wiley, 2013. <http://UCM.ebib.com/patron/FullRecord.aspx?p=1204741>.
- Lee, George, Jimmy Lin, Chuang Liu, Andrew Lorek, and Dmitriy Ryaboy. "The Unified Logging Infrastructure for Data Analytics at Twitter." *Proceedings of the VLDB Endowment* 5, no. 12 (2012): 1771–1780.
- Lee, Kyungyong, Roy Indrajit, and Vanish Talwar. "Comparing Pattern Mining on a Billion Records with HP Vertica and Hadoop," n.d. <http://www.vertica.com/2013/04/08/comparing-pattern-mining-on-a-billion-records-with-hp-vertica-and-hadoop/>.
- Leetaru, Kalev. "GDELT Data," n.d. <http://eventdata.psu.edu/data.dir/GDELT.html#.UWMC3-UnquE.twitter>.
- Lemoine, Nathan. "Python Complements R's Shortcomings." *R-bloggers*, 21:01:12. <http://www.r-bloggers.com/python-compliments-rs-shortcomings/>.
- Lengler, Ralph, and Martin Eppler. "A Periodic Table of Visualization Methods," 12:27:19. [http://www.visual-literacy.org/periodic\\_table/periodic\\_table.html](http://www.visual-literacy.org/periodic_table/periodic_table.html).
- Leo, Simone, and Gianluigi Zanetti. "Pydoop: a Python MapReduce and HDFS API for Hadoop." 819–825. HPDC '10. New York, NY, USA: ACM, 2010. doi:10.1145/1851476.1851594.
- Leonhardi, David. "Soft Skills: The 'Killer App' for Analytics," 21:29:33. <http://www.analytics-magazine.org/november-december-2011/475-soft-skills-the-killer-app-for-analytics?tmpl=component&print=1&page=>.
- Liebowitz, Jay. *Big Data and Business Analytics*. Boca Raton, FL: CRC Press, 2013. <http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=562870>.
- Lin, Jimmy. *Cloud9*, 10:20:25. <https://github.com/lintool/Cloud9>.
- . *Ivory*, 17:42:43. <https://github.com/lintool/Ivory>.
- . "Mapreduce Is Good Enough? If All You Have Is a Hammer, Throw Away Everything That's Not a Nail!" *Big Data* 1, no. 1 (2013): 28–37.
- . "Monoidify! Monoids as a Design Principle for Efficient MapReduce Algorithms." *arXiv Preprint arXiv:1304.7544* (2013). <http://arxiv.org/abs/1304.7544>.
- . *WarcBase*. Accessed October 14, 2013. <https://github.com/lintool/warcbase>.
- Lin, Jimmy, and Chris Dyer. "Data-intensive Text Processing with MapReduce," 17:36:09. <http://lintool.github.io/MapReduceAlgorithms/index.html>.
- . *Data-intensive Text Processing with MapReduce*. [San Rafael, Calif.]: Morgan & Claypool Publishers, 2010. <http://www.morganclaypool.com/doi/abs/10.2200/S00274ED1V01Y201006HLT007>.



- Lin, Jimmy, and Alek Kolcz. "Large-scale Machine Learning at Twitter." 793–804, 2012.  
<http://dl.acm.org/citation.cfm?id=2213958>.
- Lin, Jimmy, and Dmitriy Ryaboy. "Scaling Big Data Mining Infrastructure: The Twitter Experience." *ACM SIGKDD Explorations Newsletter* 14, no. 2 (2013): 6–19.
- Lin, Jimmy, Dmitriy Ryaboy, and Kevin Weil. "Full-text Indexing for Optimizing Selection Operations in Large-scale Data Analytics." In *Proceedings of the Second International Workshop on MapReduce and Its Applications*, 59–66, 2011.  
<http://dl.acm.org/citation.cfm?id=1996105>.
- Lin, Liang, Vera Lychagina, Weiran Liu, Younghee Kwon, Sagar Mittal, and Michael Wong. "Tenzing a Sql Implementation on the Mapreduce Framework" (2011).  
<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.226.772>.
- Linked Data*, 12:36:26. <http://linkeddata.org/>.
- LinkedIn. *Azkaban*, 10:27:43. <http://azkaban.github.io/azkaban2/>.
- . *Camus*. Accessed October 12, 2013. <https://github.com/linkedin/camus>.
- . *Databus*, 00:42:03. <https://github.com/linkedin/databus>.
- . *DataFu*, 10:48:47. <https://github.com/linkedin/datafu>.
- . *Hopscotch*. Accessed October 17, 2013. <https://github.com/linkedin/hopscotch>.
- . *White Elephant*. Accessed October 12, 2013.  
<https://github.com/linkedin/white-elephant>.
- Linoff, Gordon. *Data Analysis Using SQL and Excel*. Indianapolis, IN: Wiley Pub, 2008.
- Liu, Bing. *Sentiment Analysis and Opinion Mining*. San Rafael, Calif: Morgan & Claypool, 2012.
- Longo, Chris. *Pig for Sublime Text*. Accessed October 12, 2013.  
<https://github.com/chrislongo/Pig>.
- Loo, Mark P.J. van der. *Learning RStudio for R Statistical Computing: Learn to Effectively Perform R Development, Statistical Analysis, and Reporting with the Most Popular R IDE*. Birmingham: Packt Publishing, 2012.
- Lorica, Ben. "Data Science Tools: Are You 'All in' or Do You 'Mix and Match'?", 14:34:10.  
<http://strata.oreilly.com/2013/03/data-science-tools-all-in-or-mix-and-match.html>.
- . "Near Realtime, Streaming, and Perpetual Analytics," 11:18:50.  
[http://strata.oreilly.com/2013/07/near-realtime-streaming-and-perpetual-analytics.html?i\\_mm\\_mid=0adcff&cmp=em-strata-na-na-newsltr\\_20130807\\_elist](http://strata.oreilly.com/2013/07/near-realtime-streaming-and-perpetual-analytics.html?i_mm_mid=0adcff&cmp=em-strata-na-na-newsltr_20130807_elist).
- . "Simpler Workflow Tools Enable the Rapid Deployment of Models," 15:21:19.  
<http://strata.oreilly.com/2013/04/workflow-tools-enable-the-rapid-deployment-of-models.html>.
- Loshin, David. *Big Data Analytics From Strategic Planning to Enterprise Integration With Tools, Techniques, Nosql, and Graph*. Morgan Kaufmann Pub, 2013.
- . "New Paradigms for High Performance Analytical Computing" (2009).  
[https://www.ndm.net/datawarehouse/pdf/High\\_Performance\\_Paradigms.pdf](https://www.ndm.net/datawarehouse/pdf/High_Performance_Paradigms.pdf).
- Loughran, Steve. "Hoya (HBase on YARN) : Application Architecture." *Hortonworks*, 21:23:55.  
<http://hortonworks.com/blog/hoya-hbase-on-yarn-application-architecture/>.
- Loukides, Mike. *The Evolution of Data Products*. O'Reilly Media, Inc., 2011.
- Lublinsky, Boris, Kevin T Smith, and Yakubovich. *Professional Hadoop Solutions*, 2013.
- LucidWorks. *LucidWorks Big Data*, 23:22:51.  
<http://docs.lucidworks.com/display/bigdata/System+Architecture>.
- Lum, P. Y., G. Singh, A. Lehman, T. Ishkanov, M. Vejdemo-Johansson, M. Alagappan, J.

- Carlsson, and G. Carlsson. “Extracting Insights from the Shape of Complex Data Using Topology.” *Scientific Reports* 3 (February 7, 2013). doi:10.1038/srep01236.
- Macfarlane, John. *Pandoc*. Accessed October 16, 2013. <http://johnmacfarlane.net/pandoc/>.
- MADlib Project. *MADlib*, 14:25:40. <https://github.com/madlib/madlib>.
- Makoto Yui. *Hivemall*. Accessed October 17, 2013. <https://github.com/myui/hivemall>.
- Malewicz, Grzegorz, Matthew H. Austern, Aart J.C Bik, James C. Dehnert, Ilan Horn, Naty Leiser, and Grzegorz Czajkowski. “Pregel: a System for Large-scale Graph Processing.” 135–146. SIGMOD ’10. New York, NY, USA: ACM, 2010. doi:10.1145/1807167.1807184.
- Malik, Jaibeer. “ElasticSearch: Text Analysis for Content Enrichment.” *Jai’s Weblog - Tech, Security & Fun...*, 15:11:06. <http://jaibeermalik.wordpress.com/2013/03/26/elasticsearch-text-analysis-for-content-enrichment/>.
- Manning, Christopher D, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. New York: Cambridge University Press, 2008.
- Manoochehri, Michael. *Data Just Right: Practical Big Data Analytics*. Harlow: Addison-Wesley, 2013.
- Manyika, James, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers. *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute, 2011.
- Marchand, Donald A., and Joe Peppard. “Why IT Fumbles Analytics,” 06:15:55. [http://hbr.org/2013/01/why-it-fumbles-analytics/ar/1?imm\\_mid=0a4eb2&#38;cmp=em-st-rata-newsletters-stratasc-video-20130403-elst](http://hbr.org/2013/01/why-it-fumbles-analytics/ar/1?imm_mid=0a4eb2&#38;cmp=em-st-rata-newsletters-stratasc-video-20130403-elst).
- Marmanis, Haralambos, and Dmitry Babenko. *Algorithms of the Intelligent Web*. Greenwich, Conn.: Manning, 2009.
- Martin, Patrick. *Dex*, n.d. <http://dexvis.com/>.
- . *DexCharts*, 20:42:14. <https://github.com/PatMartin/DexCharts>.
- Marz, Nathan. *Big Data: Principles and Best Practices of Scalable Realtime Data Systems*. [S.l.]: O’Reilly Media, 2013.
- . *Cascalog*, 13:28:36. <https://github.com/nathanmarz/cascalog>.
- . *ElephantDB*, 11:18:06. <https://github.com/nathanmarz/elephantdb>.
- . “Human Fault-tolerance.” 00:12:41. <http://strataconf.com/strata2013/public/schedule/detail/27610>.
- . *Storm*, 22:01:45. <https://github.com/nathanmarz/storm>.
- Mashape. “List of 40+ Machine Learning APIs.” *Mashape’s Voice*, 11:16:40. <http://blog.mashape.com/post/48074869493/list-of-40-machine-learning-apis>.
- Massie, Matt. “A Powerful Big Data Trio: Spark, Parquet and Avro,” 12:27:50. <http://zenfractal.com/2013/08/21/a-powerful-big-data-trio/>.
- . *Monitoring with Ganglia*. Sebastopol, CA: O’Reilly Media, 2012. <http://proquest.safaribooksonline.com/?fpi=9781449330637>.
- Matloff, Norman S. *The Art of R Programming: Tour of Statistical Software Design*. San Francisco: No Starch Press, 2011.
- Matsumoto, Yukihiro. *Ruby*, 09:58:04. <https://www.ruby-lang.org>.
- Matusevych, Sergiy, Alex Smola, and Amr Ahmed. “Hokusai-Sketching Streams in Real Time.” *arXiv Preprint arXiv:1210.4891* (2012). <http://arxiv.org/abs/1210.4891>.
- McCallum, Q. Ethan. *Bad Data Handbook*. Beijing; Sebastopol, CA: O’Reilly, 2012.

- McDaniel, Eileen, and Stephen McDaniel. *The Accidental Analyst: Show Your Data Who's Boss*. Seattle, WA: Freakalytics, LLC, 2012.
- Mcguire, Ryan. "Tools for Testing Cassandra." *DataStax*. Accessed October 17, 2013. <http://www.datastax.com/dev/blog/tools-for-testing-cassandra>.
- McKinney, Wes. "Intro to Python for Financial Data Analysis." *Technology*, June 22, 2012. <http://www.slideshare.net/wesm/intro-to-python-for-financial-data-analysis>.
- . *Python for Data Analysis*. Sebastopol, CA: O'Reilly Media, 2012. <http://proquest.safaribooksonline.com/?fpi=9781449323592>.
- Melnik, Sergey, Andrey Gubarev, Jing Jing Long, Geoffrey Romer, Shiva Shivakumar, Matt Tolton, and Theo Vassilakis. "Dremel: Interactive Analysis of Web-scale Datasets." *Proceedings of the VLDB Endowment* 3, no. 1–2 (2010): 330–339.
- Meru, Varad. "Step-by-step MapReduce Programming," 23:22:08. <http://www.orzota.com/step-by-step-mapreduce-programming/>.
- Mesosphere. *Marathon*, 21:27:42. <https://github.com/mesosphere/marathon>.
- Metamarkets. *Druid*, n.d. <http://druid.io/>.
- Microsoft. *HDInsight*, n.d. <http://www.microsoft.com/en-us/sqlserver/solutions-technologies/business-intelligence/big-data.aspx>.
- . *StreamInsight*, n.d. <http://www.microsoft.com/sqlserver/en/us/solutions-technologies/business-intelligence/streaming-data.aspx>.
- Microsoft, Yahoo, and John Langford. *Vowpal\_wabbit*, 17:40:44. [https://github.com/JohnLangford/vowpal\\_wabbit](https://github.com/JohnLangford/vowpal_wabbit).
- Miller, Cat. "Pig Eye for the SQL Guy." *Hortonworks*, 11:20:58. <http://hortonworks.com/blog/pig-eye-for-the-sql-guy/>.
- Miller, Mark. "Solr+Hadoop = Big Data Search." *Technology*, July 15, 2013. <http://www.slideshare.net/cloudera/solrhadoopbigdatasearch>.
- Miller, Thomas W. *Modeling Techniques in Predictive Analytics: Business Problems and Solutions with R*, 2014.
- Milne, Adrian. "Dispatches from the Codeface: Complex Event Processing Made Easy (using Esper)," 20:57:20. <http://corsoftlimited.blogspot.co.uk/2013/02/complex-event-processing-made-easy.html>.
- Miner, Donald. "Data Science and Hadoop." *Technology*, October 10, 2013. <http://www.slideshare.net/DonaldMiner/data-scienceandhadoop>.
- . *Pig Vs. MapReduce: When, Why, and How*, 2013. <http://vimeo.com/73211764>.
- Miner, Donald, and Adam Shook. *MapReduce Design Patterns*. Sebastopol, CA: Oreilly, 2013.
- Mishne, Gilad, Jeff Dalton, Zhenghua Li, Aneesh Sharma, and Jimmy Lin. "Fast Data in the Era of Big Data: Twitter's Real-Time Related Query Suggestion Architecture." *arXiv Preprint arXiv:1210.7350* (2012). <http://arxiv.org/abs/1210.7350>.
- MIT. *STAR: Cluster*, n.d. <http://star.mit.edu/cluster/>.
- MongoDB. *MongoDB*, n.d. <http://www.mongodb.org/>.
- Mortar Data. *Bacon-bits*. Accessed October 17, 2013. <https://github.com/mortardata/bacon-bits>.
- . *Mortar Development Framework*. Accessed October 17, 2013. <https://github.com/mortardata/mortar>.
- . *Mortar Watchtower*. Accessed October 17, 2013. <https://github.com/mortardata/watchtower>.

Movsisyan, Mher. *Chartkick.py*, 08:32:06. <https://github.com/mher/chartkick.py>.

Mozilla Foundation. *Akela*, 20:06:03. <https://github.com/mozilla-metrics/akela>.

———. *Socorro*, 12:12:13. <https://github.com/mozilla/socorro>.

Mueller, Andreas. *Scikit-learn*, 16:38:52. <https://github.com/scikit-learn/scikit-learn>.

Muise, Adam. “HBase Technical Deep Dive.” September 18, 2013.  
<http://www.slideshare.net/adammuise/2013-sept-17thughbasetechnicalintroduction>.

———. “Hive & Performance.” July 23, 2013.  
<http://www.slideshare.net/adammuise/2013-jul-23thughivetuningdeepdive>.

Murphy, Kevin P. *Machine Learning a Probabilistic Perspective*. Cambridge, Mass.: MIT Press, 2012.  
<http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=480968>.

Murphy, Maxwell. “CFOs Ignore Big Data at Their Peril.” *Wall Street Journal*, July 18, 2013.  
<http://online.wsj.com/article/SB10001424127887323419604578571684135006800.html>.

Murthy, Arun. *Apache Hadoop Yarn: Moving Beyond Mapreduce and Batch Processing with Apache*. [S.l.]: Addison-Wesley, 2013. <http://my.safaribooksonline.com/9780133441925>.

———. “Re: About the Combiner Execution,” 20:59:38.  
[http://mail-archives.apache.org/mod\\_mbox/hadoop-hdfs-user/201107.mbox/%3C374D8F3F-B8B1-499F-BEDB-BFEE3219010C@hortonworks.com%3E](http://mail-archives.apache.org/mod_mbox/hadoop-hdfs-user/201107.mbox/%3C374D8F3F-B8B1-499F-BEDB-BFEE3219010C@hortonworks.com%3E).

Nadeau, Jacques. “Technical Overview of Apache Drill.” *Technology*, August 13, 2013.  
[http://www.slideshare.net/MapRTechnologies/technical-overview-of-apache-drill-by-jac?from\\_search=1](http://www.slideshare.net/MapRTechnologies/technical-overview-of-apache-drill-by-jac?from_search=1).

Nahum, Dotan J. *Graphene*, 06:27:27. <https://github.com/jondot/graphene>.

Nandeshwar, Ashutosh. *Tableau Data Visualization Cookbook*. Packt Publishing, Limited, 2013.

Nathan, Paco. “Pattern – an Open Source Project for Migrating Predi...” *Technology*, July 29, 2013.  
<http://www.slideshare.net/pacoid/pattern-an-open-source-project-for-migrating-predictive-models-from-sas-etc-onto-hadoop>.

Natkins, Jon. “How-to: Develop CDH Applications with Maven and Eclipse,” n.d.  
<http://blog.cloudera.com/blog/2012/08/developing-cdh-applications-with-maven-and-eclipse/>.

Neo Technology. *Neo4j*, 14:31:40. <http://www.neo4j.org/>.

Netflix. *Asgard*. Accessed October 17, 2013. <https://github.com/Netflix/asgard>.

———. *Exhibitor*, 13:38:24. <https://github.com/Netflix/exhibitor>.

———. *Genie*, 18:16:09. <https://github.com/Netflix/genie>.

———. *Lipstick*, 13:38:42. <https://github.com/Netflix/Lipstick>.

———. *Priam*, 14:10:53. <https://github.com/Netflix/Priam>.

———. *SimianArmy*, 13:39:03. <https://github.com/Netflix/SimianArmy>.

Neumeyer, Leonardo, Bruce Robbins, Anish Nair, and Anand Kesari. “S4: Distributed Stream Computing Platform.” 170–177, 2010.  
[http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=5693297](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5693297).

Nick Dimiduk. “HBase for Architects.” *Business & Mgmt*, May 28, 2013.  
<http://www.slideshare.net/xfeyr/h-base-for-architectsptx>.

Noll, Machael. “Reading and Writing Avro Files from the Command Line,” n.d.  
<http://www.michael-noll.com/blog/2013/03/17/reading-and-writing-avro-files-from-the-command-line/>.

- Noris, Basilio. *MLDemos*, 13:57:09. <https://gitorious.org/mldemos/>.
- Novus. *NVD3*. Accessed October 17, 2013. <https://github.com/novus/nvd3>.
- Nullege. “Nullege: A Search Engine for Python Source Code,” n.d. <http://nullege.com/>.
- Numenta. *NuPIC*, 22:10:46. <https://github.com/numenta/nupic>.
- Nussbaum, Bruce. *Creative Intelligence: Harnessing the Power to Create, Connect, and Inspire*. New York: HarperBusiness, 2013.
- O’Malley, Owen, and Eric Hansen. “ORC File and Vectorization - Hadoop Summit 2013.” June 28, 2013.  
[http://www.slideshare.net/oom65/orc-andvectorizationhadoopsummit?ref=http://www.slideshare.net/slideshow/embed\\_code/23625246](http://www.slideshare.net/oom65/orc-andvectorizationhadoopsummit?ref=http://www.slideshare.net/slideshow/embed_code/23625246).
- O’Neil, Cathy. *Doing Data Science*. [S.l.]: O’Reilly Media, 2013.
- O’Reilly, Tim. *How Data Science Is Transforming Health Care*. Sebastopol, CA: O’Reilly Media, 2012.  
<http://oreilly.com/data/radarreports/how-data-science-is-transforming-health-care.csp>.
- Ogievetsky, Vadim. “Facet: The Recursive Approach to Visualization.” 00:11:54.  
<http://strataconf.com/strata2013/public/schedule/detail/27485>.
- Ohlhorst, Frank. *Big Data Analytics Turning Big Data into Big Money*. Hoboken, N.J.: Wiley, 2013.  
<http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=503336>.
- Open Data Group. *Augustus*, n.d. <https://code.google.com/p/augustus/>.
- Open Knowledge Foundation. *BibServer*. Accessed October 14, 2013.  
<https://github.com/okfn/bibserver>.
- . *CKAN*. Accessed October 14, 2013. <https://github.com/okfn/ckan>.
- . *Data Explorer*. Accessed October 14, 2013. <https://github.com/okfn/dataexplorer>.
- . *Data Proxy*. Accessed October 14, 2013. <https://github.com/okfn/dataproxy>.
- . *Messytables*. Accessed October 14, 2013. <https://github.com/okfn/messytables>.
- . *Recline.js*. Accessed October 14, 2013. <https://github.com/okfn/recline>.
- OpenRefine Development Team. *OpenRefine*, n.d. <https://github.com/OpenRefine/OpenRefine>.
- OpenSCG. *BigSQL*, n.d. <http://www.bigsql.org/se/>.
- Owens, Jonathan R, Jon Lentz, and Brian Femiano. *Hadoop Real-world Solutions Cookbook*. Birmingham: Packt Pub., 2013.
- Pan, Xinghao, Evan R. Sparks, and Andre Wibisono. “MLbase: Distributed Machine Learning Made Easy.” Accessed October 20, 2013.  
[http://www.cs.berkeley.edu/~kubitron/courses/cs262a-F12/projects/reports/project15\\_report.pdf](http://www.cs.berkeley.edu/~kubitron/courses/cs262a-F12/projects/reports/project15_report.pdf).
- Paradigm4. *SciDB*, n.d. <http://www.scidb.org/>.
- Patterson, Josh. *Knitting Boar*, 10:12:22. <https://github.com/jpatanooga/KnittingBoar>.
- . “Linear Regression and Metronome.” March 24, 2013.  
<http://www.slideshare.net/jpatanooga/hadoop-summit-eu-2013-parallel-linear-regression-iterativereduce-and-yarn>.
- Patterson, Pat. *Database.com-FDW*, n.d.  
<https://github.com/metadaddy-sfdc/Database.com-FDW-for-PostgreSQL>.
- Pavlo, Andrew, Erik Paulson, Alexander Rasin, Daniel J. Abadi, David J. DeWitt, Samuel Madden, and Michael Stonebraker. “A Comparison of Approaches to Large-scale Data Analysis.” In *Proceedings of the 2009 ACM SIGMOD International Conference on*

*Management of Data*, 165–178, 2009. <http://dl.acm.org/citation.cfm?id=1559865>.

Pentaho. *Pentaho Big Data Plugin*, 21:06:06. <https://github.com/pentaho/big-data-plugin>.

Perez, Antonio Santiago. *Openlayers Cookbook*. [S.l.]: Packt Publishing Limited, 2012.

Perez, Fernando. “Literate Computing and Computational Reproducibility: IPython in the Age of Data-driven Journalism,” n.d. <http://blog.fperez.org/2013/04/literate-computing-and-computational.html>.

Perkins, Jacob. “Data Recipes: Using Hadoop to Explore Chaos.” *Data Recipes*, August 12, 2013. <http://thedatachef.blogspot.com/2013/08/using-hadoop-to-explore-chaos.html>.

———. *NLTK Trainer*, 13:00:06. <https://github.com/japerk/nltk-trainer>.

———. *Python Text Processing with NLTK 2.0 Cookbook: Over 80 Practical Recipes for Using Python’s NLTK Suite of Libraries to Maximize Your Natural Language Processing Capabilities*. Birmingham [u.a.]: Packt Publ., 2010.

Perkins, Jacob A. *Sounder*, 12:59:08. <https://github.com/thedatachef/sounder>.

———. *SwineHerd*, 12:59:19. <https://github.com/thedatachef/swineherd>.

———. *Varaha*, 12:59:14. <https://github.com/thedatachef/varaha>.

Peterson, Mike. “Using Hadoop to Expand Data Warehousing.” 00:22:47. <http://strataconf.com/strata2013/public/schedule/detail/28638>.

Phillips, Judah. *Building a Digital Analytics Organization Create Value by Integrating Analytical Processes, Technology, and People into Business Operations*, 2013. <http://proquest.safaribooksonline.com/?fpi=9780133372823>.

Pidcock, Woody, and Michael Uschold. “What Are the Differences Between a Vocabulary, a Taxonomy, a Thesaurus, an Ontology, and a Meta-model?,” 19:04:58. <http://infogrid.org/trac/wiki/Reference/PidcockArticle>.

Pike, Rob, Sean Dorward, Robert Griesemer, and Sean Quinlan. “Interpreting the Data: Parallel Analysis with Sawzall” (14:28:57). <http://lambda.csail.mit.edu/~chet/papers/others/p/pike/sawzall-sciprog.pdf>.

Pinte, Didrik. “Interactive Python Graphics/visualisation with Excel,” n.d. <http://dpinte.wordpress.com/2010/03/12/interactive-python-graphicsvisualisation-with-excel/>.

Plunkett, Tom, Brian Macdonald, Bruce Nelson, Helen Sun, Mark F Hornick, Laker, Khader Mohiuddin, et al. *Oracle Big Data Handbook*, 2014.

Preferred Infrastructure, and NTT Software Innovation Center. *Jubatus*, 17:29:59. <https://github.com/jubatus/jubatus>.

Project Blacklight. *Blacklight*, 20:57:17. <https://github.com/projectblacklight/blacklight>.

Provost, Foster, and Tom Fawcett. *Data Science for Business: [what You Need to Know About Data Mining and Data-analytic Thinking]*. Sebastopol, Calif.: O’Reilly, 2013.

Python Software Foundation. *Python*, 09:17:16. <http://www.python.org/>.

QlikTech. *QlikView*, 10:55:49. <http://www.qlikview.com/>.

Quandl. *Quandl*, n.d. <http://www.quandl.com/>.

Quantcast. *QFS*, 10:07:59. <https://github.com/quantcast/qfs>.

Rabinowitz, Nick, and Nate Agrin. “Seven Dirty Secrets of Data Visualisation,” n.d. <http://www.netmagazine.com/features/seven-dirty-secrets-data-visualisation>.

Raghunathan, Balaji. *The Complete Book of Data Anonymization: From Planning to Implementation*, 2013.

Rajaraman, Anand, and Jeffrey D Ullman. *Mining of Massive Datasets*. New York, N.Y.; Cambridge: Cambridge University Press, 2012.

Ramanujam, Srivatsan. *Pymadlib*, 14:26:03. <https://pypi.python.org/pypi/pymadlib/0.1.4>.

Rao, Jun, and Sam Shah. "Data Infrastructure at LinkedIn." *Technology*, July 8, 2013. <http://www.slideshare.net/amywtang/webscale-analyticsfb-24031929>.

Raper, Simon. "Mahout for R Users." *R-bloggers*. Accessed October 10, 2013. <http://www.r-bloggers.com/mahout-for-r-users/>.

Rapid-I. *RapidMiner*, 12:24:57. <http://rapid-i.com/content/view/181/190/>.

Redhat. *Drools*, 10:02:40. <http://www.jboss.org/drools/>.

———. *JBoss Teiid*, n.d. <http://www.jboss.org/teiid/>.

Redmond, Stephen. *QlikView for Developers Cookbook*. Birmingham: Packt Publishing, 2013. <http://public.eblib.com/EBLPublic/PublicView.do?ptiID=1192682>.

Reeve, April. *Managing Data in Motion: Data Integration, Best Practice Techniques and Technologies*, 2013.

Reich, Justin. "Small Data and Big Data Should Be Best Friends." *Education Week - EdTech Researcher*, 15:44:24. [http://blogs.edweek.org/edweek/edtechresearcher/2013/03/small\\_data\\_and\\_big\\_data\\_should\\_be\\_best\\_friends.html?cmp=SOC-SHR-FB](http://blogs.edweek.org/edweek/edtechresearcher/2013/03/small_data_and_big_data_should_be_best_friends.html?cmp=SOC-SHR-FB).

Reitz, Kenneth. *Elephant*, 19:28:21. <https://github.com/kennethreitz/elephant>.

———. *Httpbin*, 18:50:35. <https://github.com/kennethreitz/httpbin>.

———. *Requests*, 18:51:47. <https://github.com/kennethreitz/requests>.

Rendgen, Sandra, Julius Wiedemann, Paolo Ciuccarelli, Richard Saul Wurman, Simon Rogers, and Nigel Holmes. *Information graphics*. Köln: Taschen, 2012.

Reshef, David, and Yakir Reshef. *MINE*, n.d. <http://www.exploredata.net/>.

Revolution Analytics. *RHadoop*, 09:19:01. <https://github.com/RevolutionAnalytics/RHadoop>.

———. *Rhbase*, 09:20:00. <https://github.com/RevolutionAnalytics/rhbase>.

———. *Rhdfs*, 09:20:19. <https://github.com/RevolutionAnalytics/rhdfs>.

———. *Rmr2*, 09:20:34. <https://github.com/RevolutionAnalytics/rmr2>.

Riccomini, Chris. "Apache Samza: LinkedIn's Real-time Stream Processing Framework." Accessed October 17, 2013. <http://engineering.linkedin.com/data-streams/apache-samza-linkedins-real-time-stream-processing-framework>.

Richardson, Leonard. *Beautiful Soup*, 06:51:03. <https://pypi.python.org/pypi/beautifulsoup4>.

Richert, Willi, and Luis Pedro Coelho. *Building Machine Learning Systems with Python*. Birmingham, UK: Packt Pub., 2013. <http://site.ebrary.com/id/10742638>.

Rios, Miguel, and Jimmy Lin. "Distilling Massive Amounts of Data into Simple Visualizations: Twitter Case Studies." In *Workshop on Social Media Visualization at ICWSM*, 2012. <http://www.aaai.org/ocs/index.php/icwsml2/paper/download/4785/5095>.

———. "Visualizing the 'Pulse' of World Cities on Twitter." In *Seventh International AAAI Conference on Weblogs and Social Media*, 2013. <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/viewPaper/6127>.

Roch, Eric. "Drive Smarter Decisions with Big Data Using Complex Event Processing." *Technology*, July 17, 2013. <http://www.slideshare.net/perficientinc/drive-smarter-decisions-with-big-data-using-complex-event-processing>.

Rodriguez, Marko A., and Peter Neubauer. *Constructions from Dots and Lines*. ArXiv e-print, June 11, 2010. <http://arxiv.org/abs/1006.2361>.

Ross, Jeanne W, David C Robertson, and Peter Weill. *Enterprise Architecture as Strategy*:

- Creating a Foundation for Business Execution*. Boston, Mass: Harvard Business School Press, 2006.
- Ross, Noam. “Faster! Higher! Stronger! -- A Guide to Speeding Up R Code for Busy People,” n.d.  
<http://www.r-bloggers.com/faster-higher-stronger-a-guide-to-speeding-up-r-code-for-busy-people/>.
- Rossant, Cyrille. *Learning IPython for Interactive Computing and Data Visualization*, 2013.
- Roy, Indrajit. “Distributed R for Big Data,” 21:19:59.  
<http://www.vertica.com/2013/02/21/presto-distributed-r-for-big-data/>.
- Roztočil, Jakub. *HTTPIe*, 19:25:33. <https://github.com/jkbr/httpie>.
- RStudio. *RStudio*, 21:41:27. <https://github.com/rstudio/rstudio>.
- . *Shiny*, 21:41:51. <https://github.com/rstudio/shiny>.
- Russell, Matthew A. *Mining the Social Web*. Sebastopol, CA: O’Reilly, 2011.
- Salesforce.com. *Phoenix*, 10:26:43. <https://github.com/forcedotcom/phoenix>.
- Sammer, Eric. *Hadoop Operations*. Sebastopol, CA: O’Reilly, 2012.
- Sanfilippo, Salvatore. *Redis*, 09:40:34. <http://redis.io/>.
- Savage, Sam L. *The Flaw of Averages: Why We Underestimate Risk in the Face of Uncertainty*. Hoboken, N.J: Wiley, 2009.
- Schmidt, Kevin. *Programming Elastic Mapreduce: Using Aws Services to Build an End-to-end Application*. [S.l.]: O’Reilly Media, 2013.
- Schrage, Michael. “The Question All Smart Visualizations Should Ask.” *HBR Blog Network - Harvard Business Review*, 11:15:19.  
<http://blogs.hbr.org/2013/03/the-question-all-smart-visualizations/>.
- Schulte, W. Roy, and K. Mani Chandy. *Event Processing : Designing IT Systems for Agile Companies*. 1 edition. McGraw-Hill, 2009.
- Schwarzkopf, Malte, Andy Konwinski, Michael Abd-El-Malek, and John Wilkes. “Omega: Flexible, Scalable Schedulers for Large Compute Clusters.” In *Proceedings of the 8th ACM European Conference on Computer Systems*, 351–364, 2013.  
<http://dl.acm.org/citation.cfm?id=2465386>.
- ScraperWiki. *Custard*, 22:55:40. <https://github.com/scraperwiki/custard>.
- . *ScraperWiki*, 22:55:16. <https://github.com/scraperwiki/scraperwiki-python>.
- Scrapy project. *Scrapely*, 15:19:13. <https://github.com/scrapy/scrapely>.
- SeatGeek. *FuzzyWuzzy*, 22:56:53. <https://github.com/seatgeek/fuzzywuzzy>.
- Sebastian-Coleman, Laura. *Measuring Data Quality for Ongoing Improvement: a Data Quality Assessment Framework*. Waltham, MA: Morgan Kaufmann, 2013.
- Sedgewick, Robert, and Kevin Daniel Wayne. *Algorithms*. Upper Saddle River, NJ: Addison-Wesley, 2011.
- Segaran, Toby. *Programming Collective Intelligence: Building Smart Web 2.0 Applications*. Beijing; Sebastopol [Calif.]: O’Reilly, 2007.
- Segaran, Toby, and Jamie Taylor. *Programming the Semantic Web*. Beijing; Sebastopol, CA: O’Reilly, 2009.
- Segel, E. “How to Tell Stories with Data (really).” Technology, November 5, 2011.  
<http://www.slideshare.net/kris77chan/edward-segel-interactivestorytelling>.
- Segel, E., and J. Heer. “Narrative Visualization: Telling Stories with Data.” *IEEE Transactions on Visualization and Computer Graphics* 16, no. 6 (2010): 1139–1148.  
doi:10.1109/TVCG.2010.179.



- Seiler, Uwe. "Introduction to the Hadoop Ecosystem with Hadoop 2.0 Aka YARN (Java..." Business & Management, September 6, 2013.  
<http://de.slideshare.net/uweseiler/introduction-to-thehadoopecosystemjavaserbiacodecentric>.
- Sematext Group. *HBaseHUT*, 13:19:05. <https://github.com/sematext/HBaseHUT>.  
———. *HBaseWD*, 13:19:38. <https://github.com/sematext/HBaseWD>.
- SenseiDB. *SenseiDB*. Accessed October 12, 2013. <https://github.com/senseidb/sensei>.
- Shah, Sam, Craig AN Soules, Gregory R. Ganger, and Brian D. Noble. "Using Provenance to Aid in Personal File Search." In *USENIX Annual Technical Conference*, 171–184, 2007.  
[http://static.usenix.org/events/usenix07/tech/full\\_papers/shah/shah\\_html/](http://static.usenix.org/events/usenix07/tech/full_papers/shah/shah_html/).
- Shanklin, Carter. "How To Perform Spatial Analytics with Hive and Hadoop." *Hortonworks*, 20:58:31.  
<http://hortonworks.com/blog/how-to-perform-spatial-analytics-with-hive-and-hadoop/>.  
———. "ORCFile in HDP 2: Better Compression, Better Performance." *Hortonworks*. Accessed October 17, 2013.  
<http://hortonworks.com/blog/orcfile-in-hdp-2-better-compression-better-performance/>.
- Shapira, Chen (Gwen). "Data Wrangling and Oracle Connectors for Hadoop." Technology, September 30, 2013.  
<http://www.slideshare.net/gwenshap/data-wrangling-and-oracle-connectors-for-hadoop>.  
———. "Scaling ETL with Hadoop." Technology, September 13, 2013.  
<http://www.slideshare.net/gwenshap/scaling-etl-with-hadoop-shapira-3>.
- Sharethrough. *Elasticity*, 21:51:38. <https://github.com/rsrifka/elasticity>.
- Sheikh, Nauman. *Implementing Analytics: a Blueprint for Design, Development, and Adoption*. [S.l.]: Morgan Kaufmann, 2013.
- Shekhar, Jayant. "Building Recommendation Platforms with Hadoop." 00:08:52.  
<http://strataconf.com/strata2013/public/schedule/detail/27475>.
- Sheng, Lee. "Fixing Classpath Ordering Issues in Hadoop." Accessed October 17, 2013.  
<http://www.kiji.org/2013/10/08/fixing-classpath-ordering-issues-in-hadoop/>.
- Shron, Max. "Data's Missing Ingredient? Rhetoric. - Strata," 07:39:42.  
<http://strata.oreilly.com/2013/04/datas-missing-ingredient-rhetoric.html>.
- Shutterstock. *Rickshaw*, 20:55:12. <https://github.com/shutterstock/rickshaw>.
- Sid Anand. "Building a Modern Website for Scale (QCon NY 2013)." Design, June 14, 2013.  
<http://www.slideshare.net/r39132/q-con-ny2013modernwebsitescalabilityfinal-22989785>.  
———. "LinkedIn Data Infrastructure Slides (Version 2)." June 21, 2012.  
<http://www.slideshare.net/r39132/linkedin-data-infrastructure-slides-version-2-13394853>.
- Siegel, Eric. *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die*. Hoboken, N.J.: Wiley, 2013.
- Sigelman, Benjamin H., Luiz Andre Barroso, Mike Burrows, Pat Stephenson, Manoj Plakal, Donald Beaver, Saul Jaspán, and Chandan Shanbhag. "Dapper, a Large-scale Distributed Systems Tracing Infrastructure." *Google Research* (2010).  
<http://sites.google.com/site/atlediadema/training-documents/Google%20Technical%20Report%20-%20dapper,%202010.pdf>.
- Sigoure, Benoit, and Viral Bacteria. *AsyncHBase*, 09:34:33.  
<https://github.com/OpenTSDB/asynchbase>.
- Silva, Francisco Javier Blanco. *Learning Scipy for Numerical and Scientific Computing*. [S.l.]: Packt Publishing Limited, 2013.

Simon, Phil. *Too Big to Ignore: The Business Case for Big Data*, 2013.

Singh, Chandeeep. "Ingesting XML in Hive Using XPath." *The Data Stack*, 11:02:02.  
<https://communities.intel.com/community/datastack/blog/2013/08/15/hadoop-tutorials-in-gesting-xml-in-hive-using-xpath>.

Sinton, Antoine. "Building an Elasticsearch Index Offline Using Hadoop Pig," n.d.  
<http://www.poudro.com/blog/building-an-elasticsearch-index-offline-using-hadoop-pig/>.

SlamData. *Labcoat*. Accessed October 17, 2013. <https://github.com/precog/labcoat-new>.  
 ———. *Precog Platform*, 16:43:01. <https://github.com/precog/platform>.

Smeets, Bart. "Creating a Business Dashboard in R," 14:20:32.  
<http://www.r-bloggers.com/creating-a-business-dashboard-in-r/>.

Smith, Justin. *Hbase-Writer*. Accessed October 14, 2013.  
<https://code.google.com/p/hbase-writer/>.

Smith, Tim. "aRrgh - A Newcomer's (angry) Guide to Data Types in R." *GitHub*, 21:32:20.  
<https://github.com/tdsmith/aRrgh>.

Smolan, Rick, and Jennifer Erwit. *The Human Face of Big Data*. Sausalito, Calif.: Against All Odds Productions, 2012.

Snowplow Analytics. *Snowplow*, 21:53:31. <https://github.com/snowplow/snowplow>.

Solid IT. "DB-Engines Ranking," 17:25:54. <http://db-engines.com/en/ranking>.

Soztutar, Enis. "Apache HBase and HDFS - Understanding Filesystem Usa..." Technology, July 9, 2013. <http://www.slideshare.net/cloudera/internals-session-5>.

Spotify. *Luigi*, 10:49:19. <https://github.com/spotify/luigi>.  
 ———. *Snakebite*, 12:13:47. <https://github.com/spotify/snakebite>.

*Spring Data: Modern Data Access for Enterprise Java*. Sebastopol, CA: O'Reilly, 2013.

SpringSource. *Spring for Apache Hadoop*, 12:14:26.  
<https://github.com/spring-projects/spring-hadoop>.

Square. *Crossfilter*, 07:29:25. <https://github.com/square/crossfilter>.

Sripati, Praveen. "Hadoop Tips: Debugging a Hadoop MapReduce Program in Eclipse," n.d.  
<http://www.thecloudavenue.com/2012/10/debugging-hadoop-mapreduce-program-in.html>.

Stado Global Development Group. *Stado*, 10:13:14. <https://launchpad.net/stado>.

Stanford Visualization Group. *Data Wrangler*, n.d. <http://vis.stanford.edu/wrangler/app/>.

Story, Rob. *Bearcart*, 08:35:13. <https://github.com/wrobstory/bearcart>.  
 ———. *Vincent*, 08:34:28. <https://github.com/wrobstory/vincent>.

Strata Conference. *Strata Conference Santa Clara 2013: Complete Video Compilation*. [Calif?].: O'Reilly Media, 2013. <http://my.safaribooksonline.com/video/databases/9781449365394>.

Strathausen, Johann Philipp. *Graph Dracula*, 17:41:23. <https://github.com/strathausen/dracula>.

Stripe. *MoSQL*, n.d. <https://github.com/stripe/mosql>.

Stucchio, Chris. "Don't Use Hadoop - Your Data Isn't That Big." Accessed October 17, 2013.  
[http://www.chrisstucchio.com/blog/2013/hadoop\\_hatred.html](http://www.chrisstucchio.com/blog/2013/hadoop_hatred.html).

Sturm, Brian W. "The Enchanted Imagination: Storytelling's Power To Entrance Listeners." *School Library Media Research* 2 (1999).  
<http://www.eric.ed.gov/ERICWebPortal/recordDetail?accno=EJ593526>.

Sumbaly, Roshan, Jay Kreps, Lei Gao, Alex Feinberg, Chinmay Soman, and Sam Shah. "Serving Large-scale Batch Computed Data with Project Voldemort." In *Proceedings of the 10th USENIX Conference on File and Storage Technologies*, 18–18, 2012.  
<http://dl.acm.org/citation.cfm?id=2208479>.

Sumbaly, Roshan, Jay Kreps, and Sam Shah. “The Big Data Ecosystem at LinkedIn.” 1125–1134, 2013. <http://dl.acm.org/citation.cfm?id=2463707>.

Sweet, Nathan. *Kryo*, 18:23:35. <http://code.google.com/p/kryo/>.

SymPy Development Team. *SymPy*, 09:26:18. <http://sympy.org>.

SYSTAP. *Bigdata*, n.d. [http://sourceforge.net/apps/mediawiki/bigdata/index.php?title=Main\\_Page](http://sourceforge.net/apps/mediawiki/bigdata/index.php?title=Main_Page).

Tableau Software. *Tableau*, 11:00:58. <http://www.tableausoftware.com>.

Talend. *Talend Big Data*, 10:11:42. <http://www.talend.com/products/big-data>.

Talwalkara, Ameet, Tim Kraska, Rean Griffith, John Duchia, Joseph Gonzalez, Denny Britza, Xinghao Pana, Virginia Smitha, Evan Sparks, and Andre Wibisono. “MLbase: A Distributed Machine Learning Wrapper.” Accessed October 20, 2013. [http://www.cs.berkeley.edu/~ameet/dmx\\_nips.pdf](http://www.cs.berkeley.edu/~ameet/dmx_nips.pdf).

Tavory, Ami. *DAGPyte*, 14:22:31. <https://pypi.python.org/pypi/DAGPyte>.

Taylor, James. *Decision Management Systems a Practical Guide to Using Business Rules and Predictive Analytics*. [Boston, Mass.]: IBM Press/Pearson Education, 2012. <http://www.books24x7.com/marc.asp?bookid=45443>.

Teetor, Paul. *R Cookbook*. 1st ed. Beijing ; Sebastopol, CA: O’Reilly, 2011.

The Gephi Consortium. *Gephi*, n.d. <https://gephi.org/>.

*The Legacy of Fischer Black*. New York: Oxford University Press, 2005.

The Miso Project. *D3.chart*, 22:17:36. <https://github.com/misoproject/d3.chart>.

———. *Dataset.js*, 22:16:24. <https://github.com/misoproject/dataset>.

The OpenTSDB Authors, Benoit Sigoure, Geoffrey Anderson, Ion Savin, and Will Moss. *OpenTSDB*, 09:32:28. <http://opentsdb.net/>.

The PostgreSQL Global Development Group. *PostgreSQL*, 11:40:33. <http://www.postgresql.org/>.

The PyData Development Team. *Pandas*, 14:26:56. <https://github.com/pydata/pandas>.

The R Foundation. *The R Project*, 09:16:32. <http://www.r-project.org/>.

The Saddle Development Team. *SADDLE*, 22:51:09. <https://github.com/saddle/saddle>.

theguardian.com. “All Our Datasets: The Complete Index,” 13:58:15. <http://www.theguardian.com/news/datablog/interactive/2013/jan/14/all-our-datasets-index>.

Thorp, Jer. “Visualization as Process, Not Output,” n.d. [http://blogs.hbr.org/cs/2013/04/visualization\\_as\\_process.html](http://blogs.hbr.org/cs/2013/04/visualization_as_process.html).

Ting, Kathleen, and Jarek Jarcec Cecho. *Apache Sqoop Cookbook*. Sebastopol, CA: O’Reilly Media, 2013.

TinkerPop. *Gremlin*, 14:30:37. <https://github.com/tinkerpop/gremlin>.

Tran, Nga, Sreenath Bodagala, and Jaimin Dave. “Designing Query Optimizers for Big Data Problems of The Future.” *Proceedings of the VLDB Endowment* 6, no. 11 (2013). <http://db.disi.unitn.eu/pages/VLDBProgram/pdf/flash/Vertica.pdf>.

Trifacta. *Triflow*, 12:56:37. <https://github.com/trifacta/triflow>.

———. *Vega*, 12:56:53. <https://github.com/trifacta/vega>.

Trudeau, Richard J, and Richard J Trudeau. *Introduction to Graph Theory*. New York: Dover Pub., 1993.

Tsvetovat, Maksim. *SnowWhite*, 15:45:52. <https://github.com/maksim2042/snowwhite>.

Tsvetovat, Maksim, and Alexander Kouznetsov. *Social Network Analysis for Startups: [finding Connections on the Social Web]*. Sebastopol, CA: O’Reilly Media, 2011.

- Tufte, Edward R. *Visual Explanations: Images and Quantities, Evidence and Narrative*. Cheshire, Conn: Graphics Press, 1997.
- Tupper, Charles. *Data Architecture: From Zen to Reality*. Burlington, MA: Morgan Kaufmann/Elsevier, 2011.
- Twimpact. *Streamdrill*. Accessed October 20, 2013. <https://streamdrill.com/>.
- Twitter. *Ambrose*, 20:22:28. <https://github.com/twitter/ambrose>.
- . *Aurora*. Accessed October 20, 2013. <https://github.com/twitter/aurora>.
- . *Chill*, 18:16:38. <https://github.com/twitter/chill>.
- . *Elephant Twin*, 20:33:56. <https://github.com/twitter/elephant-twin>.
- . *Finagle*, 09:49:04. <https://github.com/twitter/finagle>.
- . *FlockDB*. Accessed October 20, 2013. <https://github.com/twitter/flockdb>.
- . *Hadoop-LZO*, 18:20:28. <https://github.com/twitter/hadoop-lzo>.
- . *HDFS-DU*, 20:34:54. <https://github.com/twitter/hdfs-du>.
- . *hRaven*, 13:20:23. <https://github.com/twitter/hraven>.
- . *Scalding*, 18:15:14. <https://github.com/twitter/scalding>.
- . *Snowflake*, 14:09:59. <https://github.com/twitter/snowflake>.
- . *Summingbird*, 18:15:42. <https://github.com/twitter/summingbird>.
- . *Zipkin*, 12:51:39. <https://github.com/twitter/zipkin>.
- Twitter, and Cloudera. *Parquet*, 09:55:21. <http://parquet.io/>.
- Twitter, Kevin Weil, Dmitriy Ryaboy, Chuang Liu, and Florian Leibert. *Elephant Bird*, 18:23:49. <https://github.com/kevinweil/elephant-bird>.
- Typesafe. *Akka*, 09:39:54. <http://akka.io/>.
- Üngür, Osman. *Simmetrica*, 11:12:53. <https://github.com/o/simmetrica>.
- University of Athens. *madIS*, n.d. <http://code.google.com/p/madis/>.
- University of Ljubljana. *Orange*, 12:24:23. <http://orange.biolab.si/>.
- University of Waikato. *Weka 3*, 12:25:07. <http://www.cs.waikato.ac.nz/%7Eml/weka/>.
- Uruchurtu, Linda. “38 Tools For Beautiful Data Visualisations,” 12:58:31. <http://www.elisa-dbi.co.uk/blog/38-tools-for-beautiful-data-visualisations/>.
- Valiant, Leslie G. “A Bridging Model for Parallel Computation.” *Commun. ACM* 33, no. 8 (August 1990): 103–111. doi:10.1145/79173.79181.
- Vernat, Emeric. *JavaMelody*, 12:53:37. <https://code.google.com/p/javamelody/>.
- Vertica. *Vertica Avro Parser*, 17:05:31. <https://github.com/vertica/Vertica-Extension-Packages>.
- . *Vertica-Hadoop-Connector*, 20:38:30. <https://github.com/vertica/Vertica-Hadoop-Connector>.
- Verwoerd, Jos. “Data, Data, Data: Thousands of Public Data Sources,” n.d. <http://blog.bigml.com/2013/02/28/data-data-data-thousands-of-public-data-sources/>.
- Villellas, Oscar. “Faster, More Memory Efficient SQL Queries via IOPro,” n.d. <http://continuum.io/blog/iopro-pyodbc-performance>.
- Voldemort Development Team. *Voldemort*, 10:22:32. <https://github.com/voldemort/voldemort>.
- Waber, Ben. *People Analytics How Social Sensing Technology Will Transform Business and What It Tells Us About the Future of Work*. Upper Saddle River, N.J.: FT Press, 2013. <http://proquest.safaribooksonline.com/?fpi=9780133158342>.
- Wampler, Dean. “Hadoop Data Warehousing with Hive.” 00:12:36. <http://strataconf.com/strata2013/public/schedule/detail/26899>.
- Wang, R. “Ray.” “What a Big-Data Business Model Looks Like.” *Harvard Business Review*, 22:24:37. <http://blogs.hbr.org/2012/12/what-a-big-data-business-model/>.

- Warden, Pete. *Data Source Handbook*. 1st ed. Sebastopol, CA ; Beijing [China]: O'Reilly, 2011.
- White, Brandyn. *Hadoop*, 21:24:12. <https://github.com/bwhite/hadoop>.
- White, Tom. *Hadoop: The Definitive Guide*. Farnham; Sebastopol, Calif.: O'Reilly, 2012.
- Wickham, Hadley. "Advanced R Development," 12:56:23. <https://github.com/hadley/adv-r/>.
- . "Bin-summarise-smooth: A Framework for Visualising Large Data" (14:14:18). <http://www.polisci.ucla.edu/workshops/ap-workshop-papers/Hadley%20Wickham%20Paper-%2004-08-13.pdf>.
- . *Ggplot2*, 12:59:38. <http://ggplot2.org/>.
- . *Lubridate*, 12:59:04. <https://github.com/hadley/lubridate>.
- . *Plyr*, 12:59:18. <https://github.com/hadley/plyr>.
- . "The Split-apply-combine Strategy for Data Analysis." *Journal of Statistical Software* 40, no. 1 (2011): 1–29.
- William, Sam. "Building Scalable Big Data Infrastructure Using Open Source Software." 00:09:43. <http://strataconf.com/strata2013/public/schedule/detail/27419>.
- Wilson, Chris. "Binify + D3 = Gorgeous Honeycomb Maps," 19:50:23. <http://mechanicalscribe.com/notes/binify-d3-topojson-tutorial/>.
- Wilson, Greg. "Software Carpentry: A Rational Computing Process: How and Why to Fake It," 23:04:03. <http://software-carpentry.org/blog/2013/05/rational-computing-process.html>.
- Wise, Lyndsay. *Using Open Source Platforms for Business Intelligence: Avoid Pitfalls and Maximize ROI*. Waltham, MA: Morgan Kaufmann, 2012.
- Wittenauer, Allen. "Hadoop Operations at LinkedIn." March 21, 2013. <http://www.slideshare.net/allenwittenauer/2013-hadoopsummitemea?ref=http://eventifier.co/event/hadoopsummit2013/slides>.
- . "Hadoop Performance at LinkedIn." January 19, 2013. <http://www.slideshare.net/allenwittenauer/2012-lihadoopperf>.
- Wong, Dona M. *The Wall Street Journal Guide to Information Graphics: The Dos and Don'ts of Presenting Data, Facts, and Figures*. 1st ed. New York: W.W. Norton & Co, 2010.
- Wood, David. *Linked Data*. [S.l.]: O'Reilly Media, 2013. <http://www.manning.com/dwood/>.
- Wu, Lili, Roshan Sumbaly, Chris Riccomini, Gordon Koo, Hyung Jin Kim, Jay Kreps, and Sam Shah. "Avatara: OLAP for Web-scale Analytics Products." *Proc. VLDB Endow.* 5, no. 12 (August 2012): 1874–1877.
- Xavier, Amatriain, and Justin Basilico. "The Netflix Tech Blog: System Architectures for Personalization and Recommendation," 23:52:37. <http://techblog.netflix.com/2013/03/system-architectures-for.html>.
- Xin, Reynold. "An Introduction to the Berkeley Data Analytics Stack (BDAS) Featuring Spark, Spark Streaming, and Shark." 23:11:02. <http://strataconf.com/strata2013/public/schedule/detail/27438>.
- Xin, Reynold, Josh Rosen, Matei Zaharia, Michael J. Franklin, Scott Shenker, and Ion Stoica. "Shark: SQL and Rich Analytics at Scale." *arXiv Preprint arXiv:1211.6176* (2012). <http://arxiv.org/abs/1211.6176>.
- Yahoo. *Storm-yarn*, 10:10:02. <https://github.com/yahoo/storm-yarn>.
- Yang, Fangjin. "Fast, Cheap, and 98% Right: Cardinality Estimation for Big Data." *Metamarkets*, 17:50:04. <http://metamarkets.com/2012/fast-cheap-and-98-right-cardinality-estimation-for-big-data/>.
- Yau, Nathan. *Data Points Visualization That Means Something*. Hoboken, N.J.: Wiley, 2013.

- <http://lib.myilibrary.com?id=471991>.
- . *Visualize This: The FlowingData Guide to Design, Visualization, and Statistics*. Indianapolis, Ind: Wiley, 2011. <http://www.myilibrary.com?id=317758>.
- Yelp. *Mrjob*, 21:23:32. <https://github.com/Yelp/mrjob>.
- Yhat. *Pandasql*, 09:20:38. <https://github.com/yhat/pandasql>.
- Zadrozny, Peter. *Big Data Analytics Using Splunk*. [New York, N.Y.]: Apress, 2013.
- Zaharia, MATEI, MOSHARAF Chowdhury, TATHAGATA Das, ANKUR Dave, JUSTIN Ma, MURPHY McCauley, M. Franklin, SCOTT Shenker, and I. Stoica. “Fast and Interactive Analytics over Hadoop Data with Spark.” Accessed October 17, 2013. <https://www.usenix.org/system/files/login/articles/zaharia.pdf>.
- Zaitlen, Benjamin. “Wakari and Big Finance,” 22:35:42. <http://continuum.io/blog/wakari-and-big-finance>.
- Zeyliger, Philip. “Debug Servlets, or ‘HTTP Won; Use It’ - Philz,” n.d. <http://omel.ette.org/blog/2013/02/06/debug-servlets/>.
- . “Tricks for Distributed System Debugging and Diagnosis.” 00:22:43. <http://strataconf.com/strata2013/public/schedule/detail/27194>.
- Zhai, Ke, Jordan Boyd-Graber, Nima Asadi, and Mohamad L. Alkhouja. “Mr. LDA: A Flexible Large Scale Topic Modeling Package Using Variational Inference in MapReduce.” 879–888, 2012. <http://dl.acm.org/citation.cfm?id=2187955>.