**大数据计算系统讲义**

**大数据计算系统**

1. **批处理计算框架**
2. J. Dean and S Ghemawat. MapReduce: simplified data processing on large clusters. Proc. 6th Symp. on Operating System Design and Implementation (OSDI '04), San Francisco, CA, Dec, 2004.
3. Sanjay Ghemawat, Howard Gobioff, and Shun-Tak Leung, The Google File System , 19th ACM Symposium on Operating Systems Principles, October, 2003.
4. Matei Zaharia, Mosharaf Chowdhury, Tathagata Das, Ankur Dave, Justin Ma, Murphy McCauley, Michael J. Franklin, Scott Shenker, Ion Stoica, Resilient Distributed Datasets: A Fault-Tolerant Abstraction for In-Memory Cluster Computing, University of California, Berkeley, NSDI'12 Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation, 2012.
5. Mike Burrows, Google, Inc. The Chubby lock service for loosely-coupled distributed systems. OSDI '06 Proceedings of the 7th symposium on Operating systems design and implementation. Pages 335-350.
6. Vinayak R. Borkar, Michael J. Carey, Raman Grover, Nicola Onose, Rares Vernica: Hyracks: A flexible and extensible foundation for data-intensive computing. ICDE 2011: 1151-1162
7. **流计算框架**
8. Ankit Toshniwal, Siddarth Taneja, Amit Shukla, Karthikeyan Ramasamy, Jignesh M. Patel, Sanjeev Kulkarni, Jason Jackson, Krishna Gade, Maosong Fu, Jake Donham, Nikunj Bhagat, Sailesh Mittal, Dmitriy V. Ryaboy: Storm@twitter. SIGMOD Conference 2014: 147-156
9. Matei Zaharia, Tathagata Das, Haoyuan Li, Timothy Hunter, Scott Shenker, Ion Stoica: Discretized streams: fault-tolerant streaming computation at scale. SOSP 2013: 423-438
10. **图计算框架**
11. Grzegorz Malewicz, 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, SIGMOD 2010 Proceedings of the 2010 ACM SIGMOD International Conference on Management of data, pages 135-146.
12. Yingyi Bu, Vinayak R. Borkar, Jianfeng Jia, Michael J. Carey, Tyson Condie: Pregelix: Big(ger) Graph Analytics on a Dataflow Engine. PVLDB 8(2): 161-172 (2014)
13. Joseph E. Gonzalez, Reynold S. Xin, Ankur Dave, Daniel Crankshaw, Michael J. Franklin, Ion Stoica: GraphX: Graph Processing in a Distributed Dataflow Framework. OSDI 2014: 599-613
14. Bin Shao, Haixun Wang, Yatao Li: Trinity: a distributed graph engine on a memory cloud. SIGMOD Conference 2013: 505-516
15. **存储与硬件**
16. Hongzhi Wang, Feng Xiong, Jianing Li, Shengfei Shi, Jianzhong Li, Hong Gao: Data management on new processors: A survey. Parallel Computing 72: 1-13 (2018)

**大数据管理系统**

1. **数据库与并行数据库**
2. Patricia G. Selinger Morton M. Astrahan Donald D. Chamberlin Raymond A. Lorie Thomas G. Price. Access Path Selection in a Relational Database Management System.. Proc. SIGMOD Conference, 1979, 23-34.
3. David J. DeWitt Jim Gray. Parallel Database Systems: The Future of High Performance Database Systems.. Commun. ACM, 35(6), 1992, 85-98.
4. Norbert Beckmann Hans-Peter Kriegel Ralf Schneider Bernhard Seeger. The R\*-Tree: An Efficient and Robust Access Method for Points and Rectangles.. Proc. SIGMOD Conference, 1990, 322-331.
5. **NoSQL**
6. Fay Chang, Jeffrey Dean, Sanjay Ghemawat, Wilson C. Hsieh, Deborah A. Wallach Mike Burrows, Tushar Chandra, Andrew Fikes, Robert E. Gruber, Google, Inc. Bigtable: A Distributed Storage System for Structured Data, OSDI '06 Proceedings of the 7th USENIX Symposium on Operating Systems Design and Implementation - Volume 7.
7. Avinash Lakshman and Prashant Malik. Cassandra: a decentralized structured storage system. SIGOPS Oper. Syst. Rev., 44(2):35-40, 2010.
8. Zhao Sun, Hongzhi Wang, Haixun Wang, Bin Shao, Jianzhong Li: Efficient Subgraph Matching on Billion Node Graphs. PVLDB 5(9): 788-799 (2012)
9. **NewSQL**
10. Michael Stonebraker, Ariel Weisberg: The VoltDB Main Memory DBMS. IEEE Data Eng. Bull. 36(2): 21-27 (2013)
11. Robert Kallman, Hideaki Kimura, Jonathan Natkins, Andrew Pavlo, Alex Rasin, Stanley B. Zdonik, Evan P. C. Jones, Samuel Madden, Michael Stonebraker, Yang Zhang, John Hugg, Daniel J. Abadi: H-store: a high-performance, distributed main memory transaction processing system. PVLDB 1(2): 1496-1499 (2008)