

References:

- [1] Gong XQ, Jin CQ, Wang XL, Zhang R, Zhou AY. Data-Intensive science and engineering: Requirements and challenges. *Chinese Journal of Computers*, 2012,35(8):1–16 (in Chinese with English abstract).
- [2] Li JZ, Liu XM. An important aspect of big data: Data usability. *Journal of Computer Research and Development*, 2013,50(6): 1147–1162 (in Chinese with English abstract).
- [3] Qian Z, He Y, Su C, Wu ZJ, Zhu HY. TimeStream: Reliable stream computation in the cloud. In: *Proc. of the ACM European Conf. on Computer Systems*. 2013. 1–14. [doi: 10.1145/2465351.2465353]
- [4] Zhou Y, Ooi BC, Tan KL. Dynamic load management for distributed continuous query systems. In: *Proc. of the IEEE Computer Society*. 2014. 322–323. [doi: 10.1109/ICDE.2005.54]
- [5] Zhu Y, Rundensteiner EA, Heineman GT. Dynamic plan migration for continuous queries over data streams. In: *Proc. of the ACM SIGMOD Int'l Conf. on Management of Data*. Paris, 2015. 431–442. [doi: 10.1145/1007568.1007617]
- [6] Zhou YL, Ooi BC, Tan KL, Wu J. Efficient dynamic operator placement in a locally distributed continuous query system. *LNCS 4275*, 2006. 54–71. [doi: 10.1007/11914853_5]
- [7] Fernandez RC, Migliavacca M, Kalyvianaki E, Pietzuch P. Integrating scale out and fault tolerance in stream processing using operator state management. In: *Proc. of the 2013 ACM SIGMOD Int'l Conf. on Management of Data*. 2013. 725–736. [doi: 10.1145/2463676.2465282]
- [8] Jin CQ, Qian WN, Zhou AY. Analysis and management of streaming data: A survey. *Ruan Jian Xue Bao/Journal of Software*, 2004, 15(08):1172–1181 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/15/1172.htm>
- [9] Xu Y, Kostamaa P, Zhou X, Chen L. Handling data skew in parallel joins in shared-nothing systems. In: *Proc. of the ACM SIGMOD Int'l Conf. on Management of Data*. 2008. 1043–1052. [doi: 10.1145/1376616.1376720]
- [10] Vitorovic A, Elseidy M, Koch C. Load balancing and skew resilience for parallel joins. In: *Proc. of the IEEE Int'l Conf. on Data Engineering*. IEEE, 2016. 313–324. [doi: 10.1109/ICDE.2016.7498250]
- [11] Kwon YC, Balazinska M, Howe B, Rolia J. SkewTune: Mitigating skew in mapreduce applications. In: *Proc. of the ACM SIGMOD Int'l Conf. on Management of Data*. 2012. 25–36. [doi: 10.1145/2213836.2213840]
- [12] Gufler B, Augsten N, Reiser A, Kemper A. Load balancing in MapReduce based on scalable cardinality estimates. In: *Proc. of the IEEE Int'l Conf. on Data Engineering*. 2012. 522–533. [doi: 10.1109/ICDE.2012.58]
- [13] Xing Y, Zdonik S, Hwang JH. Dynamic load distribution in the Borealis stream processor. In: *Proc. of the Int'l Conf. on Data Engineering*. IEEE, 2005. 791–802. [doi: 10.1109/ICDE.2005.53]
- [14] Xing Y, Hwang JH, Cetintemel U, Zdonik S. Providing resiliency to load variations in distributed stream processing. In: *Proc. of the Int'l Conf. on Very Large Data Bases*. ACM Press, 2006. 775–786.
- [15] Abadi DJ, Ahmad Y, Balazinska M, *et al.* The design of the Borealis stream processing engine. In: *Proc. of the 2005 CIDR Conf.*, 2005. 277–289.
- [16] Fang J, Zhang R, Fu TZJ, Zhang ZJ, Zhou AY, Zhu JH. Parallel stream processing against workload skewness and variance. *arXiv preprint arXiv:1610.05121*, 2016.
- [17] Shah MA, Hellerstein JM, Chandrasekaran S, Franklin MJ. Flux: An adaptive partitioning operator for continuous query systems. In: *Proc. of the Int'l Conf. on Data Engineering*. 2003. 25–36. [doi: 10.1109/ICDE.2003.1260779]
- [18] Gedik B. Partitioning functions for stateful data parallelism in stream processing. *Int'l Journal on Very Large Data Bases*, 2014, 23(4):517–539. [doi: 10.1007/s00778-013-0335-9]
- [19] Lin Q, Ooi BC, Wang Z, Yu C. Scalable distributed stream join processing. In: *Proc. of the ACM SIGMOD Int'l Conf*. 2015. 811–825. [doi: 10.1145/2723372.2746485]
- [20] Nasir MAU, Morales GDF, Garcia-Soriano D, Kourtellis N, Serafini M. The power of both choices: Practical load balancing for distributed stream processing engines. In: *Proc. of the IEEE Int'l Conf. on Data Engineering*. IEEE, 2015. 137–148. [doi: 10.1109/ICDE.2015.7113279]
- [21] Apache storm. <http://storm.apache.org/>
- [22] Elseidy M, Elguindy A, Vitorovic A, Koch C. Scalable and adaptive online joins. *Proc. of the VLDB Endowment*, 2014,7(6): 441–452. [doi: 10.14778/2732279.2732281]

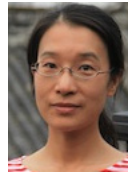
- [23] Nasir MAU, Morales GDF, Kourtellis N, Serafini M. When two choices are not enough: Balancing at scale in distributed stream processing. In: Proc. of 2016 IEEE 32nd Int'l Conf. on Data Engineering, 2016. 589–600.
- [24] Fang JH, Zhang R, Wang XT, Fu TZJ, Zhang ZJ, Zhou AY. Cost-Effective stream join algorithm on cloud system. In: Proc. of the 25th ACM Int'l on Conf. on Information and Knowledge Management. ACM Press, 2016. 1773–1782. [doi: 10.1145/2983323.2983773]
- [25] Okcan A, Riedewald M. Processing theta-joins using MapReduce. In: Proc. of the ACM SIGMOD Int'l Conf. on Management of Data. 2011. 949–960. [doi: 10.1145/1989323.1989423]
- [26] Fang JH, Wang XT, Zhang R, Zhou AY. Flexible and adaptive stream join algorithm. In: Proc. of the Asia-Pacific Web Conf. Springer Int'l Publishing, 2016. 3–16. [doi: 10.1007/978-3-319-45817-5_1]
- [27] Bruno N, Kwon YC, Wu MC. Advanced join strategies for large-scale distributed computation. Proc. of the VLDBEndowment, 2014,7(13):1484–1495. [doi: 10.14778/2733004.2733020]
- [28] The TPC-H benchmark. <http://www.tpc.org/tpch>

附中文参考文献:

- [1] 宫学庆,金澈清,王晓玲,张蓉,周傲英.数据密集型科学与工程:需求和挑战.计算机学报,2012,35(8):1–16.
- [2] 李建中,刘显敏.大数据的一个重要方面:数据可用性.计算机研究与发展,2013,50(6):1147–1162.
- [8] 金澈清,钱卫宁,周傲英.流数据分析与管理综述.软件学报,2004,15(8):1172–1181. <http://www.jos.org.cn/1000-9825/15/1172.htm>



房俊华(1985—),男,河南沈丘人,博士生,主要研究领域为分布式流处理.



张蓉(1978—),女,博士,副教授,主要研究领域为分布式计算.



王晓桐(1994—),女,硕士,主要研究领域为分布式流处理.



周傲英(1965—),男,博士,教授,博士生导师,CCF杰出会员,主要研究领域为Web数据管理,数据密集型计算,内存集群计算,分布事务处理,大数据基准测试和性能优化.