

Peer-to-peer in big data management

Bo Li · Xiaofei Liao

Received: 27 October 2013 / Accepted: 4 November 2013 / Published online: 14 November 2013
© Springer Science+Business Media New York 2013

Till now, big data is an emerging paradigm applied to massive datasets whose sizes are beyond the ability of commonly used software tools to capture, manage, and process within a tolerable elapsed time. Big data are very important for us to understand the world. Big data are often captured from various sources including social media, sensors network, scientific applications, surveillance, Internet texts and documents, business intelligence and web logs. There are several typical characteristics, including large size, heterogeneous structures, complex processing.

In big data world, there are many important techniques, including novel programming model, new system architecture, novel data storage schemes and novel data partition schemes, etc. Since P2P technologies show the good scalability, low processing cost in the last years on media streaming and file sharing services, it can be applied into the big data management to solve some key problems, including how to organize big data for indexing, searching, processing in a good distributed way, and how to assign data processing jobs in distributed environment.

The purpose of the proposed special issue is to provide a comprehensive view into recent advances in big data management. Our call for papers explicitly asked for contributions illustrating either experiences in conducting real-world

data related systems based on peer-to-peer technologies or innovative methodologies that could be useful for assessing such applications with respect to non-functional requirements, such as performance, scalability, fault tolerance, availability and security. The call for papers attracted 13 submissions. Each paper was carefully reviewed by at least three reviewers. At the end of the reviewing process, we selected six papers that we present in this special issue.

“Probabilistic Nearest Neighbor Queries of Uncertain Data via Wireless Data Broadcast” by Fangzhou Zhu et al. focuses on the performance of location-dependent query processing data for smart mobile devices. The key idea of this submission leverages the key properties of Voronoi Diagrams for Uncertain Data (UV-Diagram). This study proposes effective algorithms and demo examples to improve the performance of processing “big location-based data”.

In “DS2: a DHT-based Substrate for Distributed Services”, Lichun Li, Xin Xu, Jun Wang and Wei Wang pay their attention on the massive services deployment in IMS (IP Multimedia Subsystem) system and presents DS2, a DHT-based substrate designed for the application server farms providing distributed services. DS2 facilitates the deployment of DHT-based distributed services based on a powerful data model to manage complex data and it can support application message routing and workload migration. They have implemented the DS2 prototype and used it in ZTE’s system to enable IMS services and service routing services successfully.

“An SSD-Based Accelerator for Directory Parsing in Storage Systems Containing Massive Files” by Zhiguang Chen, Nong Xiao and Fang Liu focuses on big data storage based on SSD ((Solid State Drive) and presents an accelerator, which helps file systems to fetch the metadata of files

B. Li
Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, HongKong, China
e-mail: drlibo@gmail.com

X. Liao (✉)
School of Computer Science and Technology,
Huazhong University of Science and Technology, Hubei, China
e-mail: xfliao@hust.edu.cn

rapidly. Experimental results demonstrate that, the accelerator can speed up the directory parsing process by nearly four times compared with a file system without an accelerator.

“Handling Partitioning Skew in MapReduce using LEEN” by Shadi Ibrahim et al. focuses on the big data processing issue and presents a novel algorithm named LEEN for locality-aware and fairness-aware key partitioning in MapReduce. This submission has integrated LEEN into Hadoop and its experiments demonstrate that LEEN can efficiently achieve higher locality and reduce the amount of shuffled data. More importantly, LEEN guarantees fair distribution of the reduce inputs.

Yan Zhang et al. in “A Novel Cooperative Caching Algorithm for Massive P2P Caches” pay their attention on the massive cache data management among ISPs (Internet Service Provider). They first model the cooperative caching problem as a NP-Complete problem; then they propose a cooperative caching algorithm named cLGV (Cooperative, Lowest Global Value). The cLGV algorithm uses a new concept global value to estimate the benefits of caching or replacing an object in the cooperative caching system.

“Tag-based Personalized Image Ranking in Event Browsing” by Yeqi Lu, Yao Shen and Minyi Guo focuses on big images ranking problem and presents a new tag-based personalized image ranking algorithm in event browsing. The key idea first adopts a local matching model to assign images an original score based on whether this image satisfies users query and preference; then it proposes a global ranking model to take the local scores as initial values and make the salience scores iteratively smooth with respect to all images returned from the events of the query.

Acknowledgments The editors would like to thank all authors who submitted their work to this special issue. Last but not least we would also like to express our sincere gratitude to the following colleagues who kindly collaborated with us during the review process: Michele Amoretti, Qiang Cao, Sheng Di, Yunjun Gao, Waheed Iqbal, Victor Lee, Fuwen Liu, Haikun Liu, Yanmin Shen, LihChyun Shu, Junfeng Wang, Sheng Wang, Qingsong Wei, Chen Yu, Weihua Zhang.



Bo Li (bli@cse.ust.hk) is a professor in the Department of Computer Science and Engineering, Hong Kong University of Science and Technology. His current research interests include large-scale content distribution in the Internet, datacenter networking, cloud computing, and mobile sensor networks. He made pioneering contributions in the Internet video broadcast with a system, called Coolstreaming (the keyword had over 2,000,000 entries on Google),

which was credited as the world first large-scale Peer-to-Peer live video streaming system, which spearheaded a momentum in video broadcast industry, with no fewer than a dozen successful companies adopting the same mesh-based pull streaming technique to deliver live media content to hundreds of millions of users in the world. He has been an editor or a guest editor for a dozen of IEEE journals and magazines. He was the Co-TPC Chair for IEEE INFOCOM 2004. He received his B. Eng. Degree in the Computer Science from Tsinghua University, Beijing, and his Ph.D. degree in the Electrical and Computer Engineering from University of Massachusetts at Amherst. He is a Fellow of IEEE.



Xiaofei Liao received his Ph.D. degree in computer science and engineering from Huazhong University of Science and Technology (HUST), China, in 2005. He is now a professor in the school of Computer Science and Technology at HUST. He has served as a reviewer for many conferences and journal papers. His research interests are in the areas of system software, P2P system, cluster computing and streaming services. He is a member of the IEEE and the IEEE Computer Society.