2025春 大规模数据处理技术 Paper List

1. **FinSQL: Model-Agnostic LLMs-based Text-to-SQL Framework for Financial Analysis**, Chao Zhang, Yuren Mao, Yijiang Fan, Yu Mi, Yunjun Gao, Lu Chen, Dongfang Lou, Jinshu Lin, SIGMOD 2024

2. **Photon: A Fast Query Engine for Lakehouse Systems**, Cong Yan, Yin Lin, Yeye He, SIGMOD 2022

3. **FoundationDB: A Distributed Unbundled Transactional KeyValue Store**, Jingyu Zhou, Meng Xu, Alexander Shraer et al., SIGMOD 2021

4. **Pump Up the Volume: Processing Large Data on GPUs with Fast Interconnects**, Clemens Lutz, Sebastian Breß, Steffen Zeuch, Tilmann Rabl, Volker Markl, SIGMOD 2020

5. **SuRF: Practical Range Query Filtering with Fast Succinct Tries**, Huanchen Zhang, Hyeontaek Lim, Viktor Leis, David Andersen et al., SIGMOD 2018

6. **Scalable Link Recommendation for Influence Maximization**,Xiaolong Chen, Jing Tang, SIGKDD 2025

7. **Forward Once for All: Structural Parameterized Adaptation for Efficient Cloud-coordinated On-device Recommendation**, Kairui Fu, Zheqi Lv, Shengyu Zhang, Fan Wu, Kun Kuang, SIGKDD 2025

8. **Online Item Cold-Start Recommendation with Popularity-Aware Meta-Learning** , Yunze Luo, Yuezihan Jiang, Yinjie Jiang, Gaode Chen, Jingchi Wang, Kaigui Bian, Peiyi Li, Qi Zhang, SIGKDD 2025

9. **Collaboration of Large Language Models and Small Recommendation Models for Device-Cloud Recommendation**, Zheqi Lv, Tianyu Zhan, Wenjie Wang, Xinyu Lin, Shengyu Zhang, Wenqiao Zhang, Jiwei Li, Kun Kuang, Fei Wu, SIGKDD 2025

10. **Impoving Training Stability for Multitask Ranking Models in Recommender Systems**, Jiaxi Tang, Yoel Drori, Daryl Chang, Maheswaran Sathiamoorthy et al., SIGKDD 2023

11. **Supporting COVID-19 policy response with large-scale mobility-based modeling**, Serina Chang, Mandy L. Wilson, Bryan Lewis, Zakaria Mehrab et al., SIGKDD 2021

12. **On Sampled Metrics for Item Recommendation**, Walid Krichene, Steffen Rendle, SIGKDD 2020

13. **Temporal-Contextual Recommendation in Real-Time**, Yifei Ma, Murali Balakrishnan Narayanaswamy, Haibin Lin, Hao Ding, SIGKDD 2020

14. **Real-time Personalization using Embeddings for Search Ranking at Airbnb**, Mihajlo Grbovic, Haibin Cheng, SIGKDD 2018

15. **Accelerating Innovation Through Analogy Mining**, Tom Hope, Joel Chan, Aniket Kittur, Dafna Shahaf, SIGKDD 2017

16. **Ranking Relevance in Yahoo Search**, Dawei Yin, Yuening Hu, Jiliang Tang et al., SIGKDD 2016
17. **LLM for Data Management**, Guoliang Li, Xuanhe Zhou, xinyang zhao, VLDB 2024
18. **DOP-SQL: A General-purpose, High-utility, and Extensible Private SQL System**, Jianzhe Yu, Wei Dong, Juanru FANG, Dajun Sun, Ke Yi, VLDB 2024
19. **Auto-Tables: Synthesizing Multi-Step Transformations to Relationalize Tables without Using Examples**, Peng Li, Yeye He, Cong Yan, Yue Wang, Surajit Chaudhuri, VLDB 2023
20. **DBSP: Automatic Incremental View Maintenance for Rich Query Languages**, Mihai Budiu, Tej Chajed, Frank McSherry, Leonid Ryzhyk, Val Tannen, VLDB 2023
21. **Kora: A Cloud-Native Event Streaming Platform For Kafka**, Anna Povzner, Prince Mahajan, Jason Gustafson, Kamal Gupta et al., VLDB 2023
22. **HET: Scaling out Huge Embedding Model Training via Cache-enabled Distributed Framework**, Xupeng Miao, Hailin Zhang, Yining Shi, Xiaonan Nie et al., VLDB 2022
23. **Share the Tensor Tea: How Databases can Leverage the Machine Learning Ecosystem**, Yuki Asada, Victor Fu, Apurva Gandhi, Advitya Gemawat et al., VLDB 2022
24. **Sancus: Staleness-Aware Communication-Avoiding Full-Graph Decentralized Training in Large-Scale Graph Neural Networks**, Jingshu Peng, Zhao Chen, Yingxia Shao et al., VLDB 2022
25. **AutoExecutor: Predictive Parallelism for Spark SQL Queries**, Rathijit Sen, Abhishek Roy, Alekh Jindal et al., VLDB 2021
26. **The Ubiquity of Large Graphs and Surprising Challenges of Graph Processing: Extended Survey**, Siddhartha Sahu, Amine Mhedhbi, Semih Salihoglu et al., VLDB 2018
27. **epiC: an Extensible and Scalable System for Processing Big Data**, Dawei Jiang, Gang Chen, Beng Chin Ooi, Kian-Lee Tan, Sai Wu, VLDB 2014
28. **Incremental Learning for Multi-Interest Sequential Recommendation**, Zhikai Wang, Yanyan Shen, ICDE 2023
29. **Social Influence-based Group Representation Learning for Group Recommendation**, Hongzhi Yin, Qinyong Wang, Kai Zheng et al., ICDE 2019
30. **Faster, Higher, Stronger: Redesigning Spreadsheets for Scale**, Mangesh Bendre, Tana Wattanawaroon, Sajjadur Rahman et al., ICDE 2019