1 State of the art

1.1 Related work on data integration and cloud services

In [11], a cloud-based data management and integration system called Fusion Tables is presented. It enables data sharing, integration and collaboration between different and multiple users. Users, that can be non-IT experts, can visualize and manipulate their data in the web in a easy way. The system enables users to (i) uploading of data files from different formats; (ii) visualizing the data in different ways; (iii) integrating data from different sources belonging to multiple users. The integration process consists in a join between tables; (iv) sharing and controlling data in at levels; and (v) interacting with data in a web interface or through an API. The authors described the design foundations of Fusion Tables (such as integration with the web, easy of use, incentives for sharing and facilitate collaboration) and some examples of applications that can take advantages from the system.

The summaries for data integration are the ones I used in the first version of the related works on our paper [6, 9, 16, 18, 20]. I have no summaries of the data integration approaches I read before (I recently started doing the summaries), but this is the complete list I have read [1, 2, 3, 4, 5, 7, 8, 12, 13, 15, 17, 19, 21]. Considering these papers, most of them are frameworks/systems for data integration. Excepting the articles concerning the Google fusion tables [10], none of the works until now presented (clearly) how they integrate data (there is only a superficial description of the approach). The works focus on data quality aspects such as cost, privacy, protection and security of their integration approach. In [14], a theoretical perspective of data integration is presented, focusing on aspects such as modeling data integration applications, inconsistencies between sources, reasoning on queries and query rewriting.

References

[1] Sattam Alsubaiee, Alexander Behm, Raman Grover, Rares Vernica, Vinayak Borkar, Michael J. Carey, and Chen Li. Asterix: Scalable warehouse-style web data integration. In *Proceedings of the Ninth International Workshop on Information Integration on the Web*, IIWeb '12, pages 2:1–2:4, New York, NY, USA, 2012. ACM.

[2] Mohammed Abdullatif ALzain and E Pardede. Using Multi Shares for Ensuring Privacy in Database-as-a-Service. In *2011 44th Hawaii International Conference on System Sciences*, pages 1–9. IEEE, January 2011.

[3] S. Benkner, C. Borckholder, M. Bubak, Y. Kaniovskyi, R. Knight, M. Koehler, S. Koulouzis, P. Nowakowski, and S. Wood. A Cloud-Based Framework for Collaborative Data Management in the VPH-Share Project. In *2013 27th International Conference on Advanced Information Networking and Applications Workshops*, pages 1203–1210. IEEE, March 2013.

[4] Bin Lu and Wei Song. Research on heterogeneous data integration for Smart Grid. In *2010 3rd International Conference on Computer Science and Information Technology*, volume 3, pages 52–56. IEEE, July 2010.

[5] Reinhard Braumandl. *Quality of Service and Optimization in Data Integration Systems*. PhD thesis, University of Passau, 2002.

[6] Gianluca Correndo, Manuel Salvadores, Ian Millard, Hugh Glaser, and Nigel Shadbolt. SPARQL query rewriting for implementing data integration over linked data. In *Proceedings of the 1st International Workshop on Data Semantics - DataSem '10*, page 1, New York, New York, USA, March 2010. ACM Press.

[7] Christopher Duffy, Lorne Leonard, Gopal Bhatt, Xuan Yu, and Lee Giles. Watershed Reanalysis: Towards a National Strategy for Model-Data Integration. In *2011 IEEE Seventh International Conference on e-Science Workshops*, pages 61–65. IEEE, December 2011.

[8] Schahram Dustdar, Reinhard Pichler, Vadim Savenkov, and Hong-Linh Truong. Quality-aware service-oriented data integration: Requirements, state of the art and open challenges. *SIGMOD Rec.*, 41(1):11–19, April 2012.

[9] Ghada ElSheikh, Mustafa Y. ElNainay, Saleh ElShehaby, and Mohamed S. Abougabal. SODIM: Service Oriented Data Integration based on MapReduce. *Alexandria Engineering Journal*, 52(3):313–318, September 2013.

[10] Hector Gonzalez, Alon Halevy, Christian S. Jensen, Anno Langen, Jayant Madhavan, Rebecca Shapley, and Warren Shen. Google fusion tables: Data management, integration and collaboration in the cloud. In *Proceedings of the 1st ACM Symposium on Cloud Computing*, SoCC '10, pages 175–180, New York, NY, USA, 2010. ACM.

[11] Hector Gonzalez, Alon Y. Halevy, Christian S. Jensen, Anno Langen, Jayant Madhavan, Rebecca Shapley, Warren Shen, and Jonathan Goldberg-Kidon. Google fusion tables: Web-centered data management and collaboration. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data*, SIGMOD '10, pages 1061–1066, New York, NY, USA, 2010. ACM.

[12] Xin Hong and ChunMing Rong. Multiple Data Integration Service. In *2014 28th International Conference on Advanced Information Networking and Applications Workshops*, pages 860–865. IEEE, May 2014.

[13] Herald Kllapi, Dimitris Bilidas, Ian Horrocks, Yannis E. Ioannidis, Ernesto Jiménez-Ruiz, Evgeny Kharlamov, Manolis Koubarakis, and Dmitriy Zheleznyakov. Distributed query processing on the cloud: the optique point of view (short paper). In Mariano Rodriguez-Muro, Simon Jupp, and Kavitha Srinivas, editors, *Proceedings of the 10th International Workshop on OWL: Experiences and Directions (OWLED 2013) co-located with 10th Extended Semantic Web Conference (ESWC 2013), Montpellier, France, May 26-27, 2013.*, volume 1080 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2013.

[14] Maurizio Lenzerini. Data integration: A theoretical perspective. In *Proceedings of the Twenty-first ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, PODS '02, pages 233–246, New York, NY, USA, 2002. ACM.

[15] Asfia Mubeen, Mohd Murtuza Ahmed Khan, and Sana Mubeen Zubedi. Web Service Integration Using Cloud Data Store, 2012.

[16] Tiezheng Nie, Guangqi Wang, Derong Shen, Meifang Li, and Ge Yu. Sla-based data integration on database grids. In *Computer Software and Applications Conference, 2007. COMPSAC 2007. 31st Annual International*, volume 2, pages 613–618, July 2007.

[17] Andreas Thor and Erhard Rahm. Cloudfuice: A flexible cloud-based data integration system. In Sören Auer, Oscar Díaz, and GeorgeA. Papadopoulos, editors, *Web Engineering*, volume 6757 of *Lecture Notes in Computer Science*, pages 304–318. Springer Berlin Heidelberg, 2011.

[18] Yuan Tian, Biao Song, Jimupimg Park, and Eui-Nam Huh. Inter-cloud data integration system considering privacy and cost. In Jeng-Shyang Pan, Shyi-Ming Chen, and NgocThanh Nguyen, editors, *Computational Collective Intelligence. Technologies and Applications*, volume 6421 of *Lecture Notes in Computer Science*, pages 195–204. Springer Berlin Heidelberg, 2010.

[19] Yanxia Wang. Research on web data integration framework based on cloud computing. In *2012 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet)*, pages 2823–2826. IEEE, April 2012.

[20] S.S. Yau and Yin Yin. A privacy preserving repository for data integration across data sharing services. *Services Computing, IEEE Transactions on*, 1(3):130–140, July 2008.

[21] Peng Zhang, Yanbo Han, Zhuofeng Zhao, and Guiling Wang. Cost Optimization of Cloud-Based Data Integration System. In *2012 Ninth Web Information Systems and Applications Conference*, pages 183–188. IEEE, November 2012.