

Büyük Veri Mimarisi ve Modelleme Araçları

Bölüm

9

Dr. Öğr. Üyesi Murat TAŞYÜREK

Kayseri Üniversitesi Mühendislik Mimarlık ve Tasarım Fakültesi Bilgisayar Mühendisliği Bölümü
murattasyurek@kayseri.edu.tr

Bu Ünite Neler Öğreneceksiniz?

- ◆ Bu üniteyi tamamladıktan sonra;
- ◆ Büyük veri mimarisi için bulut sanallaştırma teknolojilerini, büyük veri ekosistemi mimarisini ve araçlarını,
- ◆ Veri toplama mekanizmasını,
- ◆ Veri akış mimarisini,
- ◆ Veri kalitesi analizini, veri modelleme işlemini açıklayabilecek,
- ◆ Hadoop kurulumunu yapabilecek ve Hadoop ile bir büyük veri örneği çözümleyebileceksiniz.

Hedefler

- ◆ Büyük veri mimarisini tanımlamak
- ◆ Büyük veri toplama mekanizmasını tanımlamak
- ◆ Veri akış mimarisini ve veri modelleme kavramlarını tanımlamak.





3. Veri akış mimarisini tanımlamak

Akış verileri, genellikle yüksek hacimlerde ve yüksek hızda sürekli olarak üretilen verileri ifade etmektedir. Veri akış mimarisi, birden çok kaynaktan büyük hacimli akış verilerini almak ve işlemek için oluşturulmuş bir yazılım bileşenleri kütüphanesidir. Geleneksel veri çözümler yöntemleri, toplu halde veri yazmaya ve okumaya odaklanırken veri akış mimarisi verileri üretildikleri anda hemen kullanır, depolamaya devam eder ve kullanım duruma göre ek bileşenleri içerebilir.

4. Veri modelleme kavramlarını tanımlamak

Veri modelleme belirli biçimsel yöntemler uygulayarak bir bilgi sistemi için bir veri modeli oluşturma sürecidir. Veri modelleme, verileri kavramsal olarak temsil etmenize ve veri nesnelere ile kurallar arasındaki ilişkiye izin verir. Veri modelleri geleneksel olarak bir projenin analiz ve tasarım aşamalarında oluşturulur. Proje gereksinimlerinin tam olarak anlaşıldığından ve karşılanacağından emin olunmasını sağlar. Veri modelleme yazılım geliştiricileri tarafından kod yazmaya başlamadan önce hataların belirlenmesine veya gerekli değişiklikleri yapılmasına öncülük ettiğinden dolayı faydalıdır.

KAYNAKLAR

1. Wu X, Zhu X, Wu G Q, et al. Data mining with big data. *IEEE transactions on knowledge and data engineering*. 2013; 26(1):97-107.
2. Gupta U G and Gupta A. Vision: a missing key dimension in the 5V Big Data framework. *Journal of International Business Research and Marketing*. 2016; 1(3):50-56.
3. Hofmann E. Big data and supply chain decisions: the impact of volume, variety and velocity properties on the bullwhip effect. *International Journal of Production Research*. 2017; 55(17):5108-5126.
4. Bhimani J, Yang Z, Leaser M and Mi N. Accelerating big data applications using lightweight virtualization framework on enterprise cloud. In 2017 IEEE High Performance Extreme Computing Conference (HPEC). 2017; 1-7.
5. Radchenko G I, Alaasam A B and Tchernykh A N. Comparative analysis of virtualization methods in big data processing. *Supercomputing Frontiers and Innovations*. 2019; 6(1):48-79.
6. Zhou G, Wang X, Chen W, Li X and Chen Z. Realization and application of geological cloud platform. *Big Earth Data*. 2020; 1-15.
7. Shilo S, Rossman H and Segal E. Axes of a revolution: challenges and promises of big data in healthcare. *Nature medicine*. 2020; 26(1):29-38.
8. Yasuoka R, Fukai T and Shinagawa T. Toward On-demand Nested Virtualization for Live-Refreshing Cloud Systems. In *The European Conference on Computer Systems*. 2020.
9. Lim J T and Nieh J. Optimizing Nested Virtualization Performance Using Direct Virtual Hardware. In *Proceedings of the Twenty-Fifth International Conference on Architectural Support*



- for Programming Languages and Operating Systems. 2020; 557-574.
10. Compastié M, Badonnel R, Festor O and He R. From virtualization security issues to cloud protection opportunities: An in-depth analysis of system virtualization models. *Computers & Security*. 2020; 97:101905.
 11. Demchenko Y, De Laat C and Membrey P. Defining architecture components of the Big Data Ecosystem. In 2014 International Conference on Collaboration Technologies and Systems (CTS). 2014; 104-112.
 12. Bourne P E, Lorsch J R and Green E D. Perspective: Sustaining the big-data ecosystem. *Nature*. 2015; 527(7576): S16-S17.
 13. Girgin M. Pazarlama ve Veri Analitiği. Pazarlamanın Artan Önemi. *Uluslararası Bankacılık Ekonomi ve Yönetim Araştırmaları Dergisi*. 2019; 2(2):1-29.
 14. Chen C P and Zhang C Y. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information sciences*. 2014; 275:314-347.
 15. Haiyun Z and Yizhe X. Sports performance prediction model based on integrated learning algorithm and cloud computing Hadoop platform. *Microprocessors and Microsystems*. 2020; 79:103322.
 16. D, Yan Z, Fu Y and Yao Z. A survey on network data collection. *Journal of Network and Computer Applications*, 2018; 116:9-23.
 17. Lyko K, Nitzschke M and Ngomo A C N. Big data acquisition. In *New Horizons for a Data-Driven Economy*, Springer. 2016; 39-61.
 18. McNeill D. IBM and the visual formation of smart cities. *Smart Urbanism: Utopian vision or false dawn*. 2016; 34-51.
 19. Ularu E G, Puican F C, Apostu A and Velicanu M. Perspectives on big data and big data analytics. *Database Systems Journal*. 2012; 3(4):3-14.
 20. Rabkin A and Katz R. Chukwa: A system for reliable large-scale log collection. In *Proceedings of LISA'10: 24th Large Installation System Administration Conference*. 2010; 163.
 21. Hoque S and Miranskyy A. Architecture for analysis of streaming data. In 2018 IEEE International Conference on Cloud Engineering (IC2E). 2018; 263-269.
 22. Mohammadi M, Al-Fuqaha A, Sorour S and Guizani M. Deep learning for IoT big data and streaming analytics: A survey. *IEEE Communications Surveys & Tutorials*. 2018; 20(4):2923-2960.
 23. Zhang F, Cao J, Khan S U, Li K and Hwang K. A task-level adaptive MapReduce framework for real-time streaming data in healthcare applications. *Future generation computer systems*. 2015; 43:149-160.
 24. Sun G, Song Y, Gong Z, Zhou X, Zhou X and Bi Y. Survey on streaming data computing system. In *Proceedings of the ACM Turing Celebration Conference-China*. 2019; 1-8.
 25. Wang R Y, Storey V C and Firth C P. A framework for analysis of data quality research. *IEEE transactions on knowledge and data engineering*. 1995; 7(4):623-640.
 26. Loshin D. Evaluating the business impacts of poor data quality. *Information Quality Journal*. 2011.
 27. Cai L and Zhu Y. The challenges of data quality and data quality assessment in the big data era. *Data science journal*. 2015; 14.
 28. Biem A, Bouillet E, Feng H, Ranganathan A, Riabov A, Verscheure O, ... and Moran C. Ibm infosphere streams for scalable, real-time, intelligent transportation services. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*. 2010; 1093-1104.
 29. Navathe S B. Evolution of data modeling for databases. *Communications of the ACM*.



1992; 35(9):112-123.

30. Goodchild M F. Geographical data modeling. *Computers & Geosciences*. 1992; 18(4):401-408.
31. Borgida A, Brachman R J, McGuinness D L and Resnick L A. CLASSIC: A structural data model for objects. *ACM Sigmod record*. 1989; 18(2):58-67.
32. Kim Y G and March S T. Comparing data modeling formalisms. *Communications of the ACM*. 1995; 38(6):103-115.
33. Gökteş B, Ömer R Ö, Duran M, Şakar S, Yılmaz M, Güler S, ... and Özdemir G. Türkiye’de Sağlık Bilgi Sistemleri Üzerine Bir Araştırma. *Ankara Sağlık Bilimleri Dergisi*, 2017; 6(1):125-138.