ifis
Institut für Informationssysteme
Technische Universität Braunschweig

Distributed Data Management

Winter Semester 2025/2026

Dr. Florian Plötzky

In this 14-week course, we discuss the foundations of distributed data management, viz., foundations of distributed databases, peer-to-peer systems, NoSQL/NewSQL systems, and modern cloud databases. We focus on a broad spectrum of concepts to paint the "big picture" and include important historical aspects along the way. Specifically we cover the following topics:

1 Foundations of Distributed Databases

In the first four lectures, we focus on the transition from classic relational databases to distributed databases and their architectures, such as mediator and federated databases. We discuss how data can be partitioned using vertical and horizontal partitioning (i.e., sharding), as well as different sharing architectures (including shared-disk and share-nothing architectures), and data allocation based on replication rules, using a practical example system in Lecture 2. In Lecture 3, we cover the basics of distributed query processing with respect to data localization and response time models. Finally, in Lecture 4, we discuss distributed transaction management, including the 2 Phase and 3 Phase Commit Protocols, distributed 2 Phase Locking, and showcase SAP HANA to illustrate different aspects of transaction processing on a real-world database.

2 Peer-to-Peer Systems

While the first part of the course focuses on distributed databases that retain all the transaction guarantees relational databases provide, we focus on the opposite case and discuss peer-to-peer (P2P) systems in the second part. The first lecture in this part introduces the foundations of P2P networks in terms of their structure as overlay networks and in terms of query mechanisms (flooding) and their limits. As a showcase, we illustrate how Bitcoin's underlying P2P architecture works and how this architecture enables a Byzantine Agreement (also covered in more detail in the third part of the course). The second lecture introduces Distributed Hash Tables (DHTs) to construct structured P2P networks. In the third lecture of this part, we discuss different network models, including Erdős-Rényi random graphs, Watts-Strogatz graphs for small-world networks, and Barabási-Albert graphs for scale-free networks. The last two lectures of this part discuss content provisioning (including BitTorrent, Kademlia, and the InterPlanetary File System) and durability aspects of structured P2P networks, i.e., load balancing and replication in DHTs, including a showcase of LOCKSS and CLOCKSS.

3 Cloud-Age Distributed Systems

The final part of the course bridges the first two parts and introduces modern distributed data processing and cloud databases. In the first lecture, we discuss the CAP and PACELC theorems along with BASE transactions and different consistency levels. Furthermore, we introduce Byzantine agreements for finding consensus in environments with potentially malicious nodes and classic consensus algorithms for fault-tolerant replication, viz., Paxos and Raft. The second and third lectures introduce NoSQL and NewSQL systems, respectively, and showcase different system families, including Amazon Dynamo (and DynamoDB), Google Bigtable and successors, Google Spanner, and CockroachDB. In the fourth lecture, we focus on cloud computing in conjunction with databases and showcase Snowflake as a cloud analytical database. The final lecture is dedicated to distributed data processing.

4 Literature

The course is based in part on textbooks but for the most part on research papers. Some research papers are explained in depth, some only cursory referenced, and some are reading hints and optional readings. The whole bibliography is depicted in the following.

Abadi, Daniel (2012). "Consistency tradeoffs in modern distributed database system design: CAP is only part of the story". In: *Computer* 45.2, pp. 37–42.

Barabási, Albert-László (2014). *Linked: How everything is connected to everything else and what it means for business, science, and everyday life*. Basic books.

Barabási, Albert-László and Réka Albert (1999). "Emergence of scaling in random networks". In: *Science* 286.5439, pp. 509–512.

Berman, Piotr, Juan Garay, Kenneth Perry, et al. (1989). "Towards optimal distributed consensus". In: *FOCS*. Vol. 89, pp. 410–415.

Brewer, Eric (2000). "Towards Robust Distributed Systems Distributed Systems". In: *Keynote at the ACM Symposium of Distributed Computing (PODC)*. URL: `https://pld.cs.luc.edu/courses/353/spr11/notes/brewer_keynote.pdf`.

Byers, John, Jeffrey Considine, and Michael Mitzenmacher (2003). "Simple Load Balancing for Distributed Hash Tables". In: *International Workshop on Peer-to-Peer Systems*. Springer, pp. 80–87.

Chang, Fay et al. (2008). "Bigtable: A distributed storage system for structured data". In: *ACM Transactions on Computer Systems (TOCS)* 26.2, pp. 1–26.

Cooper, Brian et al. (2019). "PNUTS to Sherpa: Lessons from Yahoo!'s Cloud Database". In: *Proc. of the VLDB Endowment (PVLDB)* 12.12, pp. 2300–2307.

Corbett, James et al. (2012). "Spanner: Google's Globally-Distributed Database". In: *10th USENIX Symposium on Operating Systems Design and Implementation (OSDI)*. USENIX Association, pp. 251–264.

Dageville, Benoit et al. (2016). "The Snowflake Elastic Data Warehouse". In: *Proc. of the International Conference on Management of Data (SIGMOD)*. ACM, pp. 215–226.

DeCandia, Giuseppe et al. (2007). "Dynamo: Amazon's highly available key-value store". In: *Proc. of the ACM Symposium on Operating Systems Principles (SOSP)*. Vol. 41. 6. ACM New York, NY, USA, pp. 205–220.

Elhemali, Mostafa et al. (2022). “Amazon DynamoDB: A Scalable, Predictably Performant, and Fully Managed NoSQL Database Service”. In: *Proc. of the 2022 USENIX Annual Technical Conference (USENIX-ATC*. USENIX Association, pp. 1037–1048.

Erdős, Pál and Alfréd Rényi (1959). “On the evolution of random graphs”. In: *A Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei* 5.1-2, pp. 17–61.

Färber, Franz et al. (2012). “SAP HANA Database – Data Management for Modern Business Applications”. In: *ACM Sigmod Record* 40.4, pp. 45–51.

Ghemawat, Sanjay, Howard Gobioff, and Shun-Tak Leung (2003). “The Google File System”. In: *Proc. of the ACM Symposium on Operating Systems Principles (SOSP)*, pp. 29–43.

Gilbert, Edgar N (1959). “Random graphs”. In: *The Annals of Mathematical Statistics* 30.4, pp. 1141–1144.

Gilbert, Seth and Nancy Lynch (2002). “Brewer’s conjecture and the feasibility of consistent, available, partition-tolerant web services”. In: *ACM Sigact News* 33.2, pp. 51–59.

Kadambi, Sudarshan et al. (2011). “Where in the World is My Data”. In: *Proc. of the VLDB Endowment (PVLDB)* 4.11, pp. 1040–1050.

Kleinberg, Jon M (2000). “Navigation in a small world”. In: *Nature* 406.6798, pp. 845–845.

Kleppmann, Martin (2015). “A Critique of the CAP Theorem”. In: *arXiv:1509.05393*.

– (2017). *Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems*. O’Reilly.

Kossmann, Donald (2000). “The state of the Art in Distributed Query Processing”. In: *ACM Computing Surveys (CSUR)* 32.4, pp. 422–469.

Kwak, Haewoon et al. (2010). “What is Twitter, a social network or a news media?” In: *Proc. of the International Conference on World Wide Web (WWW)*. ACM, pp. 591–600.

Lamport, Leslie (1998). “The part-time parliament”. In: *ACM Transactions on Computer Systems*.

Lee, Juchang et al. (2013). “High-Performance Transaction Processing in SAP HANA”. In: *IEEE Data Eng. Bull.* 36.2, pp. 28–33.

Leskovec, Jure and Eric Horvitz (2008). “Worldwide Buzz: Planetary-Scale Views on an Instant-Messaging Network”. In: *Proc. of the International Conference on World Wide Web (WWW)*. ACM.

Mahlmann, Peter and Christian Schindelhauer (2007). *P2P Netzwerke*. Springer.

Maniatis, Petros et al. (2005). “The LOCKSS peer-to-peer digital preservation system”. In: *ACM Transactions on Computer Systems (TOCS)* 23.1, pp. 2–50.

Manku, Gurmeet Singh and Mayank Bawa (2003). “Symphony: Distributed hashing in a small world”. In: *4th USENIX Symposium on Internet Technologies and Systems (USITS)*.

Maymounkov, Petar and David Mazieres (2002). “Kademlia: A peer-to-peer information system based on the xor metric”. In: *International workshop on peer-to-peer systems*. Springer, pp. 53–65.

Nakamoto, Satoshi (2008). *Bitcoin: A Peer-to-Peer Electronic Cash System*. URL: `https://bitcoin.org/bitcoin.pdf`.

Ongaro, Diego and John Ousterhout (2014). “In search of an understandable consensus algorithm”. In: *2014 USENIX annual technical conference (USENIX ATC)*, pp. 305–319.

Özsu, Tamer and Patrick Valduriez (2011). *Principles of Distributed Database Systems*. Springer.

Pavlo, Andrew and Matthew Aslett (2016). “What’s really new with NewSQL?” In: *ACM Sigmod Record* 45.2, pp. 45–55.

Rao, Ananth et al. (2003). “Load Balancing in Structured P2P Systems”. In: *International Workshop on Peer-to-Peer Systems*. Springer, pp. 68–79.

Ratnasamy, Sylvia et al. (2001). “A scalable content-addressable network”. In: *Proc. of the 2001 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM)*, pp. 161–172.

Rowstron, Antony and Peter Druschel (2001). “Pastry: Scalable, decentralized object location, and routing for large-scale peer-to-peer systems”. In: *IFIP/ACM International Conference on Distributed Systems Platforms and Open Distributed Processing*. Springer, pp. 329–350.

Sikka, Vishal et al. (2012). “Efficient Transaction Processing in SAP HANA Database – The End of a Column Store Myth. Categories and Subject Descriptors”. In: *Proc. of the 2012 ACM SIGMOD International Conference on Management of Data (SIGMOD)*. ACM, pp. 731–742.

Steinmetz, Ralf and Klaus Wehrle (2005). *Peer-to-Peer Systems and Applications*. Springer.

Stoica, Ion et al. (2001). “Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications”. In: *Proc. of the 2001 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM)*. ACM.

Stonebraker, Michael and Andrew Pavlo (2024). “What Goes Around Comes Around... And Around...” In: *ACM Sigmod Record* 53.2, pp. 21–37.

Stuart Holmes Rosenthal, David (2014). “Architectural choices in LOCKSS networks”. In: *Library Hi Tech* 32.1, pp. 2–10.

Taft, Rebecca et al. (2020). “CockroachDB: The Resilient Geo-Distributed SQL Database”. In: *Proc. of the International Conference on Management of Data (SIGMOD)*. ACM, pp. 1493–1509.

Terry, Douglas (2011). *Replicated Data Consistency Explained Through Baseball*. Tech. rep. Microsoft Research.

Trautwein, Dennis et al. (2022). “Design and Evaluation of IPFS: A Storage Layer for the Decentralized Web”. In: *Proc. of the ACM SIGCOMM Conference*, pp. 739–752.

Watts, Duncan and Steven Strogatz (1998). “Collective dynamics of ‘small-world’networks”. In: *Nature* 393.6684, pp. 440–442.

Wingerath, Wolfgang, Felix Gessert, and Norbert Ritter (2019). “NoSQL & Real-Time Data Management in Research and Practice”. In: *Tutorial @ BTW*.