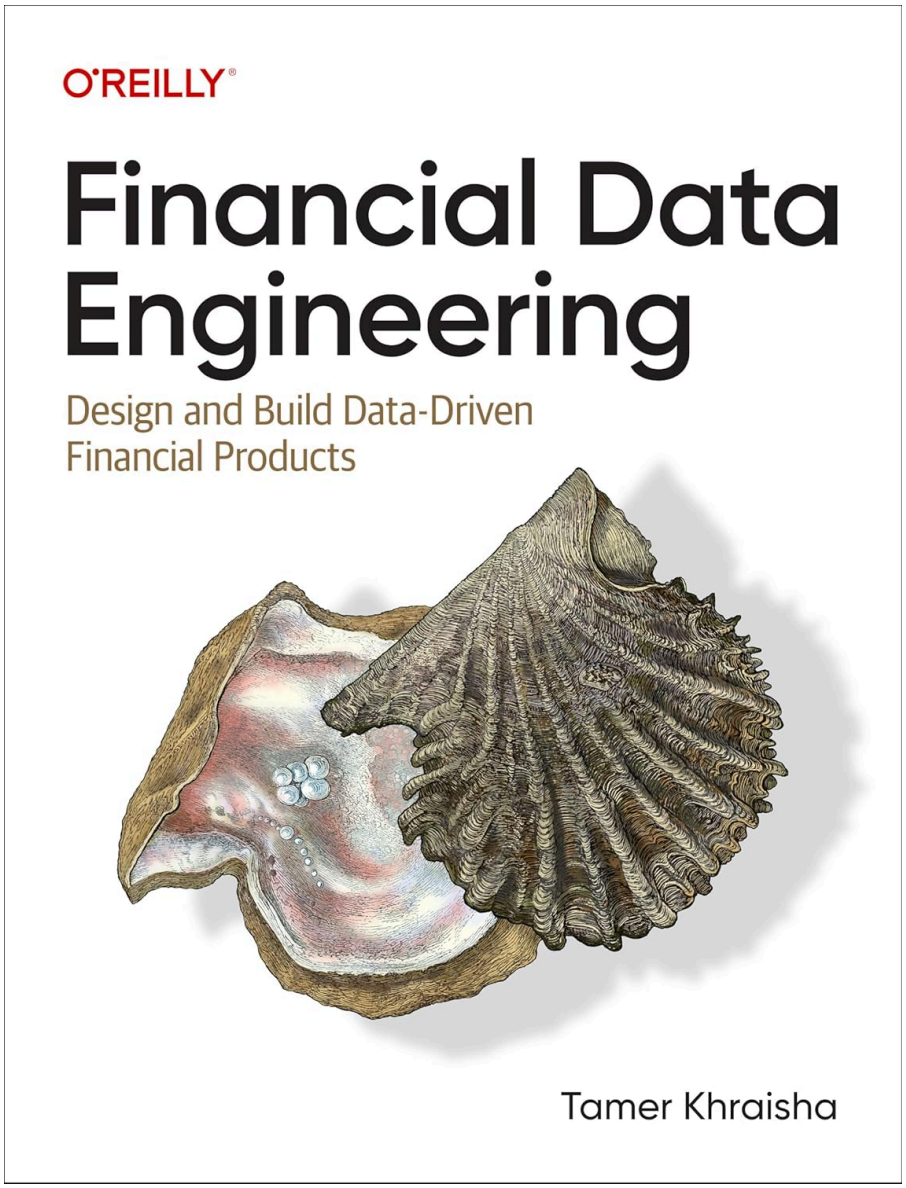

Book References

This file contains a comprehensive list of references that were used in writing the book "*Financial Data Engineering: Design and Build Data-Driven Financial Products*" by O'Reilly Media

Chapter 1

- Sironi, Paolo. FinTech innovation: from robo-advisors to goal based investing and gamification. John Wiley & Sons, 2016.
- Brose, Margarita S., Mark D. Flood, Dilip Krishna, and Bill Nichols, eds. Handbook of Financial Data and Risk Information II: Software and Data. Cambridge University Press, 2014.
- Armour, John, Daniel Awrey, Paul Lyndon Davies, Luca Enriques, Jeffrey Neil Gordon, Colin P. Mayer, and Jennifer Payne. Principles of financial regulation. Oxford University Press, 2016.
- Lee, In. "Big data: Dimensions, evolution, impacts, and challenges." Business Horizons 60, no. 3 (2017): 293-303.
- Fang, Bin, and Peng Zhang. "Big data in finance." Big data concepts, theories, and applications (2016): 391-412.
- Gai, Prasanna. Systemic risk: the dynamics of modern financial systems. OUP Oxford, 2013.
- Soldatos, John, and Dimosthenis Kyriazis. Big Data and Artificial Intelligence in Digital Finance: Increasing Personalization and Trust in Digital Finance Using Big Data and AI. Springer Nature, 2022.
- Bodie, Zvi, Alex Kane, and Alan J. Marcus. Investments (SIE). McGraw-Hill Education, 2014.
- Unifying Market Data: Consolidated Tape Providers in the EU & US, by Christopher Buttigieg and Nathan Fenech
- Francis, Jack Clark, and Dongcheol Kim. Modern portfolio theory: Foundations, analysis, and new developments. John Wiley & Sons, 2013.
- Hull, John C. Options, Futures, and Other Derivatives, Global Edition. Pearson, 2021.
- Bessis, Joel. Risk management in banking. John Wiley & Sons, 2015.
- De Prado, Marcos Lopez. Advances in financial machine learning. John Wiley & Sons, 2018.

- Dixon, Matthew F., Igor Halperin, and Paul Bilokon. Machine learning in finance. Vol. 1170. Berlin/Heidelberg, Germany: Springer International Publishing, 2020.
- de Prado, Marcos M. López. Machine learning for asset managers. Cambridge University Press, 2020.
- Li, Yinheng, Shaofei Wang, Han Ding, and Hang Chen. "Large language models in finance: A survey." In Proceedings of the Fourth ACM International Conference on AI in Finance, pp. 374-382. 2023.
- Hilpisch, Yves. Python for finance: mastering data-driven finance. O'Reilly Media, 2018.
- Guida, Tony. Big data and machine learning in quantitative investment. John Wiley & Sons, 2019.
- Karasan, Abdullah. Machine Learning for Financial Risk Management with Python. " O'Reilly Media, Inc.", 2021.
- Coqueret, Guillaume, and Tony Guida. Machine Learning for Factor Investing: Python Version. CRC Press, 2023.
- Hasan, Md Morshadul, József Popp, and Judit Oláh. Current landscape and influence of big data on finance. Journal of Big Data 7, no. 1 (2020): 1-17.
- "The Top Data Management Challenges Faced by Financial Firms" By John Eley
- Capital Markets Infrastructure: An Industry Reinventing Itself" By McKinsey
- Tagliani, Matthew. The practical guide to Wall Street: equities and derivatives. Vol. 479. John Wiley & Sons, 2009.
- Big data: Profitability, potential, and problems in banking. The Financial Brand Retrieved from
- Using domain analysis to model microservices, by Microsoft

Chapter 2

Journal Articles

- Khraisha, Tamer. "A Holistic Approach to Financial Data Science: Data, Technology, and Analytics." The Journal of Financial Data Science (2020).

- Goodhart, Charles AE, and Maureen O'Hara. "High frequency data in financial markets: Issues and applications." Journal of Empirical Finance 4, no. 2-3 (1997): 73-114.

- Brooks, Chris, Andreas GF Hoepner, David McMillan, Andrew Vivian, and Chardin Wese Simen. "Financial data science: the birth of a new financial research paradigm complementing econometrics?." The European Journal of Finance 25, no. 17 (2019): 1627-1636.

- How to read a financial report: wringing vital signs out of the numbers by John A. Tracy and Tage C. Tracy

- Loughran, Tim, and Bill McDonald. "Textual analysis in finance." Annual Review of Financial Economics 12 (2020): 357-375.

- Hu, Gang, Koren M. Jo, Yi Alex Wang, and Jing Xie. "Institutional trading and Abel Noser data." *Journal of Corporate Finance* 52 (2018): 143-167.

- Purcell, Darren. "Standardisation put to the test: How reference data standards have evolved in an era of unprecedented regulatory and technological change." Journal of Securities Operations & Custody 11, no. 4 (2019): 327-336.

- Kivelä, Mikko, Alex Arenas, Marc Barthelemy, James P. Gleeson, Yamir Moreno, and Mason A. Porter. "Multilayer networks." Journal of Complex Networks 2, no. 3 (2014): 203-271.

- Hautsch, Nikolaus, and Winfried Pohlmeier. "Econometric analysis of financial transaction data: pitfalls and opportunities." (2001).

Online articles

- "Data everywhere, all at once: data management challenges for sell-side firms" By EDT

- "Reference Data Poses Significant Challenges for the Capital Markets, Study from Accenture and Greenwich Associates Shows" By Accenture

- "Bloomberg vs. Capital IQ vs. Factset vs. Refinitiv" By WallStreetPrep

- "Big Data in The Banking Industry: The Main Challenges and Use Cases" by Alexey Shalimov

- "Reference Data and its Role in Operational Risk Management" By Capgemini

Books

- Brose, Margarita S., Mark D. Flood, Dilip Krishna, and Bill Nichols, eds. Handbook of Financial Data and Risk Information I. Vol. 1. Cambridge University Press, 2014.
- Koop, Gary. *Analysis of financial data*. John Wiley & Sons Inc., 2022.
- Guida, Tony. Big data and machine learning in quantitative investment. John Wiley & Sons, 2019.
- Kaufman, Perry J. *Trading Systems and Methods,+ Website*. Vol. 591. John Wiley & Sons, 2013.
- Denev, Alexander, and Saeed Amen. *The Book of Alternative Data: A Guide for Investors, Traders and Risk Managers*. John Wiley & Sons, 2020.
- Groot, Martijn. *A primer in financial data management*. Academic Press, 2017.
- Pauley, Justin. Unlocking financial data: a practical guide to technology for equity and fixed income analysts. " O'Reilly Media, Inc.", 2017.
- Andersen, T. G., Davis, R. A., Kreiß, J. P., & Mikosch, T. V. (Eds.). (2009). Handbook of financial time series. Springer Science & Business Media.
- Ruppert, David, and David S. Matteson. Statistics and data analysis for financial engineering. Vol. 13. New York: Springer, 2011.
- Taylor, Stephen J. *Modelling financial time series*. world scientific, 2008.
- Fabozzi, Frank J., ed. *Encyclopedia of Financial Models, Volume I*. John Wiley & Sons, 2012.
- Harris, Larry. Trading and exchanges: Market microstructure for practitioners. OUP USA, 2003.
- Fridson, Martin S., and Fernando Alvarez. Financial statement analysis: a practitioner's guide. John Wiley & Sons, 2022.
- Köseoğlu, Sinem Derindere, ed. *Financial Data Analytics: Theory and Application*. Springer Nature, 2022.
- Aldridge, Irene, and Marco Avellaneda. *Big data science in finance*. John Wiley & Sons, 2021.

- Bouchaud, Jean-Philippe, Julius Bonart, Jonathan Donier, and Martin Gould. *Trades, quotes and prices: financial markets under the microscope*. Cambridge University Press, 2018.

- De Prado, Marcos Lopez. *Advances in financial machine learning*. John Wiley & Sons, 2018.

- Mitra, Gautam, and Leela Mitra, eds. *The handbook of news analytics in finance*. John Wiley & Sons, 2011.

- Hafez, Peter, Ricard Matas Navarro, Inna Grinis, Francisco Gomez, Marko Kangrga, and Alan Liu. "Factor investing with sentiment: a look at asia-pacific markets." *Available at SSRN 4323549* (2020).

- Cavanillas, José María, Edward Curry, and Wolfgang Wahlster. New horizons for a data-driven economy: a roadmap for usage and exploitation of big data in Europe. Springer Nature, 2016.

- Mention, Anne-laure, and Dimitrios G. Salampasis, eds. Transformation Dynamics in FinTech: An Open Innovation Ecosystem Outlook. Vol. 7. World Scientific, 2021.

- Ang, Clifford S., and Clifford S. Ang. Analyzing financial data and implementing financial models using R. Basel: Springer, 2015.

- Berman, Karen, and Joe Knight. Financial intelligence, revised edition: A manager's guide to knowing what the numbers really mean. Harvard Business Review Press, 2013.

Research papers that include sample data from CRSP

- Brogaard, Jonathan, Dan Li, and Ying Xia. Stock liquidity and default risk. *Journal of Financial Economics* 124, no. 3 (2017): 486-502.
- Jegadeesh, N., & Kim, W. (2010). Do analysts herd? An analysis of recommendations and market reactions. The Review of Financial Studies, 23(2), 901-937.
- Gandhi, Priyank, and Hanno Lustig. Size anomalies in US bank stock returns. The Journal of Finance 70, no. 2 (2015): 733-768.

Research papers that include sample data from Compustat Global

- De Jong, A., Kabir, R., & Nguyen, T. T. (2008). Capital structure around the world: The roles of firm-and country-specific determinants. Journal of Banking & Finance, 32(9), 1954-1969.
- Brown, James R., Gustav Martinsson, and Bruce C. Petersen. Law, stock markets, and innovation. The Journal of Finance 68, no. 4 (2013): 1517-1549.

- Erkens, David H., Mingyi Hung, and Pedro Matos. Corporate governance in the 2007–2008 financial crisis: Evidence from financial institutions worldwide. Journal of corporate finance 18, no. 2 (2012): 389-411.

Research papers that include sample data from TAQ

- Bessembinder, Hendrik. Quote-based competition and trade execution costs in NYSE-listed stocks. Journal of Financial Economics 70, no. 3 (2003): 385-422.
- Hendershott, Terrence, Charles M. Jones, and Albert J. Menkveld. Does algorithmic trading improve liquidity?. The Journal of finance 66, no. 1 (2011): 1-33.
- Hvidkjaer, Soeren. A trade-based analysis of momentum. The Review of Financial Studies 19, no. 2 (2006): 457-491.

Research papers that include sample data from I/B/E/S

- Kumar, Alok. "Who gambles in the stock market?." The journal of finance 64, no. 4 (2009): 1889-1933.
- Malmendier, Ulrike, and Devin Shanthikumar. "Do security analysts speak in two tongues?." The Review of Financial Studies 27, no. 5 (2014): 1287-1322.
- De Bondt, Werner FM, and Richard H. Thaler. "Do security analysts overreact?." The American economic review (1990): 52-57.

Research papers that include sample data from OptionMetrics

- Kelly, Bryan, Ľuboš Pástor, and Pietro Veronesi. The price of political uncertainty: Theory and evidence from the option market. The Journal of Finance 71, no. 5 (2016): 2417-2480.
- González-Urteaga, Ana, and Gonzalo Rubio. The cross-sectional variation of volatility risk premia. Journal of Financial Economics 119, no. 2 (2016): 353-370.
- Jensen, Mads Vestergaard, and Lasse Heje Pedersen. Early option exercise: Never say never. Journal of Financial Economics 121, no. 2 (2016): 278-299.

Research papers that include sample data from TRACE

- Tsai, Hui-Ju, and Yangru Wu. Bond and stock market response to unexpected dividend changes. Journal of Empirical Finance 30 (2015): 1-15.
- Becker, Bo, and Victoria Ivashina. Reaching for yield in the bond market. The Journal of Finance 70, no. 5 (2015): 1863-1902.
- Cici, Gjergji, Scott Gibson, and John J. Merrick Jr. Missing the marks? Dispersion in corporate bond valuations across mutual funds. Journal of Financial Economics 101, no. 1 (2011): 206-226.

Research papers that include sample data from ORBIS

- Tian, Gloria Y., and Garry Twite. Corporate governance, external market discipline and firm productivity. Journal of Corporate Finance 17, no. 3 (2011): 403-417.

- Bloom, Nicholas, Aprajit Mahajan, David McKenzie, and John Roberts. Why do firms in developing countries have low productivity?. American Economic Review 100, no. 2 (2010): 619-623.
- Succurro, Marianna. Financial bankruptcy across European countries. International Journal of Economics and Finance 9, no. 7 (2017): 132-146.

Research papers that include sample data from SDC Platinum

- Jackson, Andrew R. Trade generation, reputation, and sell-side analysts. The journal of finance 60, no. 2 (2005): 673-717.
- Tang, Dragon Yongjun, and Yupu Zhang. Do shareholders benefit from green bonds?. Journal of Corporate Finance 61 (2020): 101427.
- Pathan, Shams. Strong boards, CEO power and bank risk-taking. Journal of banking & finance 33, no. 7 (2009): 1340-1350.

Chapter 3

- Berman, Jules J. Principles and practice of big data: preparing, sharing, and analyzing complex information. Academic Press, 2018.
- Baker, Tim. The Challenges of Symbology in Financial Data. IEX Cloud. 2021.
- Allocation rules for the financial instrument global identifier (FIGI) standard, by OpenFigi
- Official ANNA page for ISO Financial Identifiers, by ANNA
- CUSIP Global Services Identifiers, by CUSIP
- A Unique Transaction Identifier for securities: All you need to know, by SWIFT
- BIS(editors): "Committee on Payments and Market Infrastructures Board of the International Organization of Securities Commissions Technical Guidance, Harmonisation of the Unique Transaction Identifier." February, 2017.
- Wikipedia's page for Refinitiv Identification Code
- OCC Option Symbology Initiative (OCI) Implementation Guide. Interactivebrokers.com. September 2012. Retrieved 23 November 2017.
- What's in a name? The Bloomberg Global ID is reborn as the FIGI, by Peter Warms
- Open Market Data Initiative, by Bloomberg 2012

- CIRCULAR CSSF 08/365 on Alternative Instrument Identifier, by CSSF
- Demystifying cryptocurrency and digital assets, by PWC
- Digital Assets: Let's Talk about Tokens: https://www.six-group.com/en/blog/2023/what-is-a-token.html
- Pearce, Esther, and Milo O. Pererson. History of the standard industrial classification. 1957. Available at
- Business Identifier Code (BIC), by SWIFT
- Common language for digital identifiers, by ISO
- Five things you didn't know about global payments, by ISO
- Financial Instrument Identifiers: One international standard setting the bar, by Stephan Dreyer
- Securities Identifiers, by prospectus.com
- Financial Instrument Identification Market Practice, by Securities Market Practice Group (SMPG)

Chapter 4

- Rao, Delip, Paul McNamee, and Mark Dredze. "Entity linking: Finding extracted entities in a knowledge base." Multi-source, multilingual information extraction and summarization (2013): 93-115.
- Kanungsukkasem, Nont, Ponrudee Netisopakul, and Teerapong Leelanupab. "Recognition of NASDAQ stock symbols in Tweets." In 2014 6th International Conference on Knowledge and Smart Technology (KST), pp. 12-16. IEEE, 2014.
- Hardeniya, Nitin, Jacob Perkins, Deepti Chopra, Nisheeth Joshi, and Iti Mathur. Natural language processing: python and NLTK. Packt Publishing Ltd, 2016.
- Herzog, Thomas N., Fritz J. Scheuren, and William E. Winkler. Data quality and record linkage techniques. Vol. 1. New York: Springer, 2007.
- Christen, P. Data matching: concepts and techniques for record linkage, entity resolution, and duplicate detection. Springer, 2012.

- Mitra, Gautam, and Leela Mitra, eds. The handbook of news analytics in finance. John Wiley & Sons, 2011.

- Using NER to detect relevant entities in finance, by RavenPack

- Sharnagat, Rahul. "Named entity recognition: A literature survey." Center For Indian Language Technology (2014): 1-27.

- Binette, Olivier, and Rebecca C. Steorts. "(Almost) all of entity resolution." Science Advances 8, no. 12 (2022). https://www.science.org/doi/full/10.1126/sciadv.abi8021

- Papadakis, George, Ekaterini Ioannou, Emanouil Thanos, and Themis Palpanas. The four generations of entity resolution. San Rafael, CA: Morgan & Claypool Publishers, 2021.

- Vatsalan, Dinusha, Peter Christen, and Vassilios S. Verykios. "A taxonomy of privacy-preserving record linkage techniques." Information Systems 38, no. 6 (2013): 946-969.

- Li, Lingli, Jianzhong Li, and Hong Gao. "Rule-based method for entity resolution." IEEE Transactions on Knowledge and Data Engineering 27, no. 1 (2014): 250-263.

- Whang, Steven Euijong, David Marmaros, and Hector Garcia-Molina. "Pay-as-you-go entity resolution." IEEE Transactions on Knowledge and Data Engineering 25, no. 5 (2012): 1111-1124.

- Li, J., Sun, A., Han, J., & Li, C. (2020). A survey on deep learning for named entity recognition. IEEE Transactions on Knowledge and Data Engineering, 34(1), 50-70.

- Zhu, Chenguang. Machine Reading Comprehension: Algorithms and Practice. Elsevier, 2021.

- Ignatow, Gabe, and Rada Mihalcea. An introduction to text mining: Research design, data collection, and analysis. Sage Publications, 2017.

- Rau, L. F. (1991, January). Extracting company names from text. In Proceedings the Seventh IEEE Conference on Artificial Intelligence Application (pp. 29-30). IEEE Computer Society.

- Named Entity Recognition using Deep Learning(ELMo Embedding+ Bi-LSTM), by Subham Sarkar

- Liu, X., Chen, H., & Xia, W. (2022). Overview of named entity recognition. Journal of Contemporary Educational Research, 6(5), 65-68.

- Named Entity Recognition with Hugging Face Transformers: A Beginner's Guide, by Ganesh Lokare

- Using NER to detect relevant entities in finance, by RavenPack

- Flood, Mark, John Grant, Haiping Luo, Louiqa Raschid, Ian Soboroff, and Kyungjin Yoo. "Financial entity identification and information integration (feiii) challenge: the report of the organizing committee." In Proceedings of the Second International Workshop on Data Science for Macro-Modeling, pp. 1-4. 2016.

- Financial Use Cases For Named Entity Recognition (NER), by Sharon Yang, CFA

- Chan, Samuel WK, and Mickey WC Chong. "Sentiment analysis in financial texts." Decision Support Systems 94 (2017): 53-64.

- Christophides, Vassilis, Vasilis Efthymiou, Themis Palpanas, George Papadakis, and Kostas Stefanidis. "An overview of end-to-end entity resolution for big data." ACM Computing Surveys (CSUR) 53, no. 6 (2020): 1-42.

- Marrero, Mónica, Julián Urbano, Sonia Sánchez-Cuadrado, Jorge Morato, and Juan Miguel Gómez-Berbís. "Named entity recognition: fallacies, challenges and opportunities." Computer Standards & Interfaces 35, no. 5 (2013): 482-489.

- Zhang, Dongsong, and Lina Zhou. "Discovering golden nuggets: data mining in financial application." IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 34, no. 4 (2004): 513-522.

- Talburt, John R. Entity resolution and information quality. Elsevier, 2011.

- Alvarado, Julio Cesar Salinas, Karin Verspoor, and Timothy Baldwin. "Domain adaption of named entity recognition to support credit risk assessment." In Proceedings of the Australasian Language Technology Association Workshop 2015, pp. 84-90. 2015.

- Lee, Meisin, Lay-Ki Soon, Eu-Gene Siew, and Ly Fie Sugianto. "CrudeOilNews: An Annotated Crude Oil News Corpus for Event Extraction." arXiv preprint arXiv:2204.03871 (2022).

- Jacobs, Gilles, and Véronique Hoste. "SENTiVENT: enabling supervised information extraction of company-specific events in economic and financial news." Language Resources and Evaluation 56, no. 1 (2022): 225-257.

- Repke, Tim, and Ralf Krestel. "Extraction and representation of financial entities from text." In Data Science for Economics and Finance: Methodologies and Applications, pp. 241-263. Cham: Springer International Publishing, 2021.

- Gschwind, Thomas, Christoph Miksovic, Julian Minder, Katsiaryna Mirylenka, and Paolo Scotton. "Fast record linkage for company entities." In 2019 IEEE International Conference on Big Data (Big Data), pp. 623-630. IEEE, 2019.

- Schild, Christopher-Johannes. "Linking'Orbis' Company Data with Establishment Data from the German Federal Employment Agency." NO. WP-GRLC-2016-02 (2016).

- Christen, Peter, and Ross Gayler. "Towards scalable real-time entity resolution using a similarity-aware inverted index approach." (2008).

- González-Carrasco, Israel, Jose Luis Jiménez-Márquez, Jose Luis López-Cuadrado, and Belén Ruiz-Mezcua. "Automatic detection of relationships between banking operations using machine learning." Information Sciences 485 (2019): 319-346.

- He, Xin, Kaiyong Zhao, and Xiaowen Chu. "AutoML: A survey of the state-of-the-art." Knowledge-Based Systems 212 (2021): 106622.

Chapter 5

- Karkošková, Soňa. "Data governance model to enhance data quality in financial institutions." Information Systems Management 40, no. 1 (2023): 90-110.

- Mahanti, Rupa, and Rupa Mahanti. Data governance and compliance. Springer Singapore, 2021.

- Liu, Grace. "Data quality problems troubling business and financial researchers: A literature review and synthetic analysis." Journal of Business & Finance Librarianship 25, no. 3-4 (2020): 315-371

- Ince, Ozgur S., and R. Burt Porter. "Individual equity return data from Thomson Datastream: Handle with care!." Journal of Financial Research 29, no. 4 (2006): 463-479

- Madnick, Stuart E., Richard Y. Wang, Yang W. Lee, and Hongwei Zhu. "Overview and framework for data and information quality research." Journal of data and information quality (JDIQ) 1, no. 1 (2009): 1-22.

- Abraham, Rene, Johannes Schneider, and Jan Vom Brocke. "Data governance: A conceptual framework, structured review, and research agenda." International journal of

information management 49 (2019): 424-438.

- Amiram, Dan, Zahn Bozanic, and Ethan Rouen. "Financial statement errors: Evidence from the distributional properties of financial statement numbers." Review of accounting studies 20 (2015): 1540-1593.

- Ladley, John. Data governance: How to design, deploy, and sustain an effective data governance program. Academic Press, 2019.

- Saleem, Danial. "Data Governance Strategies for AI/ML in Banking Applications." International Journal of Computer Science and Technology 7, no. 1 (2023): 95-117.

- Puri, Colin, Doo Soon Kim, Peter Z. Yeh, and Kunal Verma. "Implementing a data lineage tracker." In Data Warehousing and Knowledge Discovery: 14th International Conference, DaWaK 2012, Vienna, Austria, September 3-6, 2012. Proceedings 14, pp. 390-403. Springer Berlin Heidelberg, 2012.

- Gudivada, Venkat, Amy Apon, and Junhua Ding. "Data quality considerations for big data and machine learning: Going beyond data cleaning and transformations." International Journal on Advances in Software 10, no. 1 (2017): 1-20.

- Du, Jie, and Lina Zhou. "Improving financial data quality using ontologies." Decision Support Systems 54, no. 1 (2012): 76-86.

- Sadowski, Jathan. "When data is capital: Datafication, accumulation, and extraction." Big data & society 6, no. 1 (2019): 2053951718820549.

- Lo, Andrew W. "Moore's Law vs. Murphy's Law in the financial system: who's winning?." Journal of Investment Management, no.1 (2016): 17-38

- Wand, Yair, and Richard Y. Wang. "Anchoring data quality dimensions in ontological foundations." Communications of the ACM 39, no. 11 (1996): 86-95.

- Three priorities for financial institutions to drive a next-generation data governance framework, by Ernst & Young

- THE ULTIMATE DATA PRIVACY GUIDE FOR BANKS AND FINANCIAL INSTITUTIONS https://www.ngdata.com/data-privacy-guide-for-banks-and-financial-institutions/

- Data lineage: Data origination and where it moves over time, by Deloitte

- Improving investor confidence with ISO standards, by Barnaby Lewis

- Financial Data Protection: How Do Banks Backup Data?, By Vinchin
- What is encryption?, By Google
- Wing, Jeannette M. "The data life cycle." Harvard Data Science Review 1, no. 1 (2019): 6.
- Our GDPR summary for Financial Services, By PWC
- Aggarwal, Charu C., and S. Yu Philip, eds. Privacy-preserving data mining: models and algorithms. Springer Science & Business Media, 2008.
- Brownlees, Christian T., and Giampiero M. Gallo. "Financial econometric analysis at ultra-high frequency: Data handling concerns." Computational statistics & data analysis 51, no. 4 (2006): 2232-2245.
- Ruohonen, Jukka, and Sini Mickelsson. "Reflections on the Data Governance Act." Digital Society 2, no. 1 (2023): 10.
- Ultimate guide to IBOR vs ABOR and other Books of Record (BORs), by Dr Ian Hunt
- Adkins, Heather, Betsy Beyer, Paul Blankinship, Piotr Lewandowski, Ana Oprea, and Adam Stubblefield. Building secure and reliable systems: best practices for designing, implementing, and maintaining systems. O'Reilly Media, 2020.
- El Emam, Khaled, Lucy Mosquera, and Richard Hoptroff. Practical synthetic data generation: balancing privacy and the broad availability of data. O'Reilly Media, 2020.
- Piechocki, Michal. "Overview of international experiences with data standards and identifiers applicable for big data analysis." In IFC-Bank Indonesia Satellite Seminar on “Big Data” at the ISI Regional Statistics Conference. 2017.
- Kerwer, Dieter. "Governing financial markets by international standards." New Modes of Governance in the Global System: Exploring Publicness, Delegation and Inclusiveness (2006): 77-100.
- Liu, Jing, Yang Xiao, Hui Chen, Suat Ozdemir, Srinivas Dodle, and Vikas Singh. "A survey of payment card industry data security standard." IEEE Communications Surveys & Tutorials 12, no. 3 (2010): 287-303.
- Batini, Carlo, Cinzia Cappiello, Chiara Francalanci, and Andrea Maurino. "Methodologies for data quality assessment and improvement." ACM computing surveys (CSUR) 41, no. 3 (2009): 1-52.

- Hussain, Kazim, and Elsa Prieto. "Big data in the finance and insurance sectors." New horizons for a data-driven economy: A roadmap for usage and exploitation of big data in Europe (2016): 209-223.

- Gdovin, Filip. "Graphical Representation of Data Lineage in the Data Governance Tool." PhD diss., Masarykova univerzita, Fakulta informatiky, 2017.

- Martinen, Michael, George Black, Ripple Bhullar, and Victor Marranca. "Consolidated Audit Trail: Strategic planning and best practices." Journal of Securities Operations & Custody 10, no. 1 (2018): 77-83.

- Arbuckle, Luk, and Khaled El Emam. Building an anonymization pipeline: creating safe data. O'Reilly Media, 2020.

- Raghunathan, Balaji. The complete book of data anonymization: from planning to implementation. CRC Press, 2013.

- Murthy, Suntherasvaran, Asmidar Abu Bakar, Fiza Abdul Rahim, and Ramona Ramli. "A comparative study of data anonymization techniques." In 2019 IEEE 5th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing,(HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS), pp. 306-309. IEEE, 2019.

- Aggarwal, Charu C., and Philip S. Yu. "A general survey of privacy-preserving data mining models and algorithms." Privacy-preserving data mining: models and algorithms (2008): 11-52.

- 10 Misunderstandings related to anonymisation, By Agencia Española de Protección de Datos

- Anonymization and Pseudonymization Policy, by Robert Bateman

- Data Encryption in Finance, by Apricorn

- Why you can't really anonymize your data, by Pete Warden

Chapter 6

- Marinescu, Dan C. Cloud computing: theory and practice. Morgan Kaufmann, 2022.

- Munar, Antoni, Esteban Chiner, and Ignacio Sales. "A big data financial information management architecture for global banking." In 2014 international conference on future

internet of things and cloud, pp. 385-388. IEEE, 2014.

- Lemaire, Maude. Refactoring At Scale. " O'Reilly Media, Inc.", 2020.
- Newman, Sam. Monolith to microservices: evolutionary patterns to transform your monolith. O'Reilly Media, 2019.
- Talukder, Asoke K., and Lawrence Zimmerman. "Cloud economics: Principles, costs, and benefits." Cloud computing: Principles, systems and applications (2010): 343-360.
- Wu, Caesar, Rajkumar Buyya, and Kotagiri Ramamohanarao. "Cloud pricing models: Taxonomy, survey, and interdisciplinary challenges." ACM Computing Surveys (CSUR) 52, no. 6 (2019): 1-36.
- Dotson, Chris. Practical Cloud Security: A Guide for Secure Design and Deployment. " O'Reilly Media, Inc.", 2023.
- Indrasiri, Kasun, and Sriskandarajah Suhothayan. Design Patterns for Cloud Native Applications. " O'Reilly Media, Inc.", 2021.
- Isom, Pamela K., and Kerrie Holley. Is your company ready for Cloud: Choosing the best Cloud adoption strategy for your business. IBM Press, 2012.
-
- Software Development Life Cycle: Finding a Model That Works, by ADAM MURRAY
- The Minimum Elements For a Software Bill of Materials (SBOM), by The United States Department of Commerce
- Why the C programming language still rules, by Serdar Yegulalp
- What is a mainframe?, by IBM
- Dempster, Michael Alan Howarth, Juho Kanniainen, John Keane, and Erik Vynckier, eds. High-performance computing in finance: problems, methods, and solutions. CRC Press, 2018.
- Banks Tiptoe Toward Their Cloud-Based Future, by Cate Andrews
- Taulli, Tom. Modern Mainframe Development." O'Reilly Media, Inc.", 2012.
- Agile vs. waterfall project management, by Dan Radigan

- Yu, Xiaojun, and Qiaoyan Wen. "A view about cloud data security from data life cycle." In 2010 international conference on computational intelligence and software engineering, pp. 1-4. IEEE, 2010.
- Zhang, Zan, Guofang Nan, and Yong Tan. "Cloud services vs. on-premises software: competition under security risk and product customization." Information Systems Research 31, no. 3 (2020): 848-864.

- Sabiri, Khadija, and Faouzia Benabbou. "Methods migration from on-premise to cloud." IOSR Journal of Computer Engineering 17, no. 2 (2015): 58-65.

- Robey, Robert, and Yuliana Zamora. Parallel and high performance computing. Simon and Schuster, 2021.

- Tian, Xinhui, Rui Han, Lei Wang, Gang Lu, and Jianfeng Zhan. "Latency critical big data computing in finance." The Journal of Finance and Data Science 1, no. 1 (2015): 33-41.

- Frino, Alex, Vito Mollica, and Robert I. Webb. "The impact of co-location of securities exchanges' and traders' computer servers on market liquidity." Journal of Futures Markets 34, no. 1 (2014): 20-33.

- Laughlin, Gregory, Anthony Aguirre, and Joseph Grundfest. "Information transmission between financial markets in Chicago and New York." Financial Review 49, no. 2 (2014): 283-312.

- Pottathuparambil, Robin, Jack Coyne, Jeffrey Allred, William Lynch, and Vincent Natoli. "Low-latency FPGA based financial data feed handler." In 2011 IEEE 19th Annual International Symposium on Field-Programmable Custom Computing Machines, pp. 93-96. IEEE, 2011.

- Morris, Gareth W., David B. Thomas, and Wayne Luk. "FPGA accelerated low-latency market data feed processing." In 2009 17th IEEE Symposium on High Performance Interconnects, pp. 83-89. IEEE, 2009.

- Denholm, Stewart, Hiroaki Inoue, Takashi Takenaka, Tobias Becker, and Wayne Luk. "Low latency FPGA acceleration of market data feed arbitration." In 2014 IEEE 25th

Chapter 7

- Kurose, James F., and Keith W. Ross. "Computer networking: A top-down approach." Pearson (2020).

- Buchanan, Mark. "Physics in finance: Trading at the speed of light." Nature 518, no. 7538 (2015): 161-163.

- Sharma, Garima, Vikas Tripathi, and Awadhesh Srivastava. "Recent trends in big data ingestion tools: A study." In Research in Intelligent and Computing in Engineering: Select Proceedings of RICE 2020, pp. 873-881. Springer Singapore, 2021.

- Hlupić, Tomislav, and Josip Puniš. "An Overview of Current Trends in Data Ingestion and Integration." In 2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO), pp. 1265-1270. IEEE, 2021.

- Chawla, Harsh, Pankaj Khattar, Harsh Chawla, and Pankaj Khattar. "Data ingestion." Data Lake Analytics on Microsoft Azure: A Practitioner's Guide to Big Data Engineering (2020): 43-85.

- Meehan, John, Cansu Aslantas, Stan Zdonik, Nesime Tatbul, and Jiang Du. "Data Ingestion for the Connected World." In CIDR, vol. 17, pp. 8-11. 2017.

- Sawant, Nitin, Himanshu Shah, Nitin Sawant, and Himanshu Shah. "Big Data Ingestion and Streaming Patterns." Big Data Application Architecture Q & A: A Problem-Solution Approach (2013): 29-42.

- Pasetto, Davide, Karol Lynch, Robert Tucker, Brendan Maguire, Fabrizio Petrini, and Hubertus Franke. "Ultra low latency market data feed on IBM PowerENTM." Computer Science-Research and Development 26, no. 3-4 (2011): 307-315.

- International Conference on Application-Specific Systems, Architectures and Processors, pp. 36-40. IEEE, 2014.

- Roodyn, Neil, and Wolfgang Emmerich. "An architectural style for multiple real-time data feeds." In Proceedings of the 21st International Conference on Software Engineering, pp. 564-572. 1999.

- Complete Guide to Data Ingestion: Types, Process, and Best Practices, by IBM

- Secure File Transfer for Financial Services: Best Practices for MFT and Automated File Transfer, By Bob Ertl

- Data formats supported by Azure Data Explorer for ingestion, By Microsoft

- How a Data Ingestion Framework Powers Large Dataset Usage, by Snowflake

- The Difference Between Real-Time, Near Real-Time, and Batch Processing in Big Data, by Christy Wilson

- SQL Injection Attack: How It Works, Examples and Prevention, by Bar Hofesh

- Payment APIs: What Are They and How Do They Work?, By Stax
- Top Payment Gateway APIs for Developers, by Atatus
- Brodsky, Laura, and Liz Oakes. "Data sharing and open banking." McKinsey & Company 1105 (2017).
- Omarini, Anna Eugenia. "Banks and FinTechs: How to develop a digital open banking approach for the bank's future." International Business Research 11, no. 9 (2018): 23-36.
- Guibaud, Sophie. "How to develop a profitable, customer-focused digital banking strategy: Open banking services and developer-friendly APIs." Journal of Digital Banking 1, no. 1 (2016): 6-12.
- Zachariadis, Markos, and Pinar Ozcan. "The API economy and digital transformation in financial services: The case of open banking." SWIFT Institute Working Paper No. 2016-001 (2017).
- Kohli, Manu, and Edgardo Suarez. "Centralized solution to securely transfer payment information electronically to banks from multiple enterprise resource planning (ERP) systems." In 2016 International Conference on Information Technology (ICIT), pp. 275-282. IEEE, 2016.
- Zaharia, Matei, Mosharaf Chowdhury, Tathagata Das, Ankur Dave, Justin Ma, Murphy McCauly, Michael J. Franklin, Scott Shenker, and Ion Stoica. "Resilient distributed datasets: A Fault-Tolerant abstraction for In-Memory cluster computing." In 9th USENIX Symposium on Networked Systems Design and Implementation (NSDI 12), pp. 15-28. 2012.
- Bruggink, Diederik, Pierre Karsten, and Carlo de Meijer. "The European cards environment and ISO 20022." Journal of Payments Strategy & Systems 6, no. 1 (2012): 80-99.
- Schmidt, Isabel. "ISO 20022: Understanding the scale of change." Journal of Digital Banking 5, no. 2 (2020): 126-136.
- ISO 20022 for Dummies, by SWIFT

Chapter 8

- Gilbert, Seth, and Nancy Lynch. "Perspectives on the CAP Theorem." Computer 45, no. 2 (2012): 30-36.

- Özsu, M. Tamer, and Patrick Valduriez. Principles of distributed database systems. Vol. 2. Englewood Cliffs: Prentice Hall, 1999.

- Bondi, André B. "Characteristics of scalability and their impact on performance." In Proceedings of the 2nd international workshop on Software and performance, pp. 195-203. 2000.

- How to Organize your Data Lake, by Rodrigo Souza

- Giebler, Corinna, Christoph Gröger, Eva Hoos, Holger Schwarz, and Bernhard Mitschang. "Leveraging the data lake: Current state and challenges." In Big Data Analytics and Knowledge Discovery: 21st International Conference, DaWaK 2019, Linz, Austria, August 26–29, 2019, Proceedings 21, pp. 179-188. Springer International Publishing, 2019.

- Bernstein, Philip A., and Nathan Goodman. "Concurrency control in distributed database systems." ACM Computing Surveys (CSUR) 13, no. 2 (1981): 185-221

- Gorelik, Alex. The enterprise big data lake: Delivering the promise of big data and data science. O'Reilly Media, 2019.

- Giebler, Corinna, Christoph Gröger, Eva Hoos, Rebecca Eichler, Holger Schwarz, and Bernhard Mitschang. "The data lake architecture framework: a foundation for building a comprehensive data lake architecture." In Conference for Database Systems for Business, Technology and Web (BTW), vol. 70469. 2021.
- Sawadogo, Pegdwendé, and Jérôme Darmont. "On data lake architectures and metadata management." Journal of Intelligent Information Systems 56 (2021): 97-120.

- Sharma, Ben. Architecting data lakes: data management architectures for advanced business use cases. O'Reilly Media, 2018.

- How Financial Institutions Can Use AWS to Address Regulatory Reporting, by Jiwan Panjiker, Brian Tang, and Raviteja Yelamanchili

- The SQL Standard – ISO/IEC 9075:2023 (ANSI X3.135), by Brad Kelechava

- Lock contention, suspension, deadlock, and timeout, by IBM

- Normalization in DBMS, by Study Tonight

- What Is a Data Warehouse?, by Oracle
- Kdb vs TimescaleDB, by InfluxData
- Time-Series Database: An Explainer, by Timescale
- Graph Technology for Financial Services, by Neo4j
- What is Data Modeling? Conceptual, Physical, Logical, by Matthew Groves
- Van Landuyt, Dimitri, Julien Benaouda, Vincent Reniers, Ansar Rafique, and Wouter Joosen. "A Comparative Performance Evaluation of Multi-Model NoSQL Databases and Polyglot Persistence." In Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing, pp. 286-293. 2023.
- Płuciennik, Ewa, and Kamil Zgorzałek. "The multi-model databases–a review." In Beyond Databases, Architectures and Structures. Towards Efficient Solutions for Data Analysis and Knowledge Representation: 13th International Conference, BDAS 2017, Ustroń, Poland, May 30-June 2, 2017, Proceedings 13, pp. 141-152. Springer International Publishing, 2017.
- Meier, Andreas, and Michael Kaufmann. SQL & NoSQL databases. Berlin/Heidelberg, Germany: Springer Fachmedien Wiesbaden, 2019.
- Santos, Maribel Yasmina, and Carlos Costa. "Data models in NoSQL databases for big data contexts." In Data Mining and Big Data: First International Conference, DMBD 2016, Bali, Indonesia, June 25-30, 2016. Proceedings 1, pp. 475-485. Springer International Publishing, 2016.
- What is Schema-on-Read vs Schema-on-Write?, by Dremio
- What is data modeling?, by IBM
- Zheng, Zibin, Shaoan Xie, Hongning Dai, Xiangping Chen, and Huaimin Wang. "An overview of blockchain technology: Architecture, consensus, and future trends." In 2017 IEEE international congress on big data (BigData congress), pp. 557-564. Ieee, 2017.
- Chowdhury, Mohammad Jabed Morshed, Alan Colman, Muhammad Ashad Kabir, Jun Han, and Paul Sarda. "Blockchain versus database: A critical analysis." In 2018 17th IEEE International conference on trust, security and privacy in computing and communications/12th IEEE international conference on big data science and engineering (TrustCom/BigDataSE), pp. 1348-1353. IEEE, 2018.

- McConaghy, Trent, Rodolphe Marques, Andreas Müller, Dimitri De Jonghe, Troy McConaghy, Greg McMullen, Ryan Henderson, Sylvain Bellemare, and Alberto Granzotto. "Bigchaindb: a scalable blockchain database." white paper, BigChainDB (2016): 53-72.

- Raikwar, Mayank, Danilo Gligoroski, and Goran Velinov. "Trends in development of databases and blockchain." In 2020 Seventh International Conference on Software Defined Systems (SDS), pp. 177-182. IEEE, 2020.

- Lantz, Lorne, and Daniel Cawrey. Mastering blockchain. O'Reilly Media, 2020.

- Kumar, Neeraj, N. Gayathri, Md Arafatur Rahman, and B. Balamurugan, eds. Blockchain, Big Data and Machine Learning: Trends and Applications. CRC Press, 2020.

- Fernandes, Diogo, and Jorge Bernardino. "Graph Databases Comparison: AllegroGraph, ArangoDB, InfiniteGraph, Neo4J, and OrientDB." Data 10 (2018): 0006910203730380.

- Banking resilience at global scale with Distributed SQL, by Cockroach Labs

- How CockroachDB does distributed, atomic transactions, by Matt Tracy

- Parallel Commits: An atomic commit protocol for globally distributed transactions, by Nathan VanBenschoten

- How to build a payments system that scales to infinity (with examples), by Charlie Custer

- Idempotency's role in financial services (with examples), by Cassie McAllister

- Graph Data Science Use Cases: Fraud and Anomaly Detection, Neo4j white paper by Jaimie Chung

- Banking Circle: Money Laundering Beware: A Modern Approach to AML with Machine Learning and Graphs, a presentation by Ruben Menke, Lead Data Scientist at Banking Circle

- CockroachDB: The Definitive Guide: Distributed Data at Scale, by Guy Harrison, Jesse Seldess, and Ben Darnell

- Exploring the Feasibility of Blockchain for High-Storage Applications, by Anthony Clarke

Chapter 9

- Brownlees, Christian T., and Giampiero M. Gallo. "Financial econometric analysis at ultra-high frequency: Data handling concerns." Computational statistics & data analysis 51, no. 4 (2006): 2232-2245.

- Optimize query computation, by Google
- What is high performance computing?, by Google
- Dean, Jeffrey, and Luiz André Barroso. "The tail at scale." Communications of the ACM 56, no. 2 (2013): 74-80.
- For compliance and latency in banking, move the data closer to the customer, by Jessica Edwards
- Lopez de Prado, Marcos. "Supercomputing for Finance: A Gentle Introduction (Presentation Slides)." *Available at SSRN 2907803* (2017).
- DECIMAL PLACES vs SIGNIFICANT FIGURES for FX Rates: Why the wrong choice can have a major impact on your company accounts, by FXLoader

Chapter 10

- Amit Duvedi, Balaji Mohanam, Andy Still, and Andrew Ash. Financial Governance for Data Processing in the Cloud " O'Reilly Media, Inc.", 2019.
- Storment, J. R., and Mike Fuller. Cloud FinOps. " O'Reilly Media, Inc.", 2023.
- Hu, Gang, Koren M. Jo, Yi Alex Wang, and Jing Xie. "Institutional trading and Abel Noser data." Journal of Corporate Finance 52 (2018): 143-167.
- Karumuri, Suman, Franco Solleza, Stan Zdonik, and Nesime Tatbul. "Towards observability data management at scale." ACM SIGMOD Record 49, no. 4 (2021): 18-23.
- Majors, Charity, Liz Fong-Jones, and George Miranda. Observability Engineering. " O'Reilly Media, Inc.", 2022.
- Andy Petrella. Fundamentals of Data Observability" O'Reilly Media, Inc.", 2022.
- Lu, Jie, Anjin Liu, Fan Dong, Feng Gu, Joao Gama, and Guangquan Zhang. "Learning under concept drift: A review." IEEE transactions on knowledge and data engineering 31, no. 12 (2018): 2346-2363.
- Riedesel, Jamie. Software Telemetry: Reliable logging and monitoring. Simon and Schuster, 2021.
- 7 top data quality management tools, by George Lawton

- Incident management for high-velocity teams, by Atlassian

- Girling, Philippa X. Operational risk management: a complete guide for banking and fintech. John Wiley & Sons, 2022.

- Chau, Derek, and Maarten van Dijck Nemcsik. Anti-money laundering transaction monitoring systems implementation: Finding anomalies. John Wiley & Sons, 2020.

- Chapelle, Ariane. Operational risk management: Best practices in the financial services industry. John Wiley & Sons, 2019.

Chapter 11

- Building Domain Driven Microservices, by Chandra Ramalingam

- Chen, Cathy, Niall Richard Murphy, Kranti Parisa, D. Sculley, and Todd Underwood. Reliable Machine Learning. " O'Reilly Media, Inc.", 2021.

- Huyen, Chip. Designing machine learning systems. " O'Reilly Media, Inc.", 2022.

- Gorodetskaya, Olga, Yana Gobareva, and Mikhail Koroteev. "Forecasting Time Series in the Banking Sector Using a Machine Learning Pipeline." In 2021 14th International Conference Management of large-scale system development (MLSD), pp. 1-5. IEEE, 2021.

- Fang, Haokun, and Quan Qian. "Privacy preserving machine learning with homomorphic encryption and federated learning." Future Internet 13, no. 4 (2021): 94.

- What is Secure Multiparty Computation?, by Inpher

- Masters, Oliver, Hamish Hunt, Enrico Steffinlongo, Jack Crawford, Flavio Bergamaschi, Maria E. Dela Rosa, Caio C. Quini, Camila T. Alves, Feranda de Souza, and Deise G. Ferreira. "Towards a homomorphic machine learning big data pipeline for the financial services sector." Cryptology ePrint Archive (2019)

- WHAT IS A FEATURE STORE IN MACHINE LEARNING?, by Snowflake

- Domain-Driven Design Principles for Microservices, by Tomas Fernandez