Rutgers Business School

Important Course Details

Course Information

- Course Number and Course Title: 22:544:645: Big Data and Cloud Computing, cross-listed as 26:198:641 Advanced Database Systems
- Term and Year: Fall 2025
- Office Building & Room Number:
    - Livingston Campus: BRR-5066
    - Newark Campus: 1WP-1015A
- Class Building, Room Number, & Campus:
    - 22:544:645:01: 1WP-402, Newark Campus
    - 26:198:641-01: 1WP-402, Newark Campus
- Class Meeting Times:
    - 22:544:645:01: W, 10:00-12:50
    - 26:198:641-01: W, 10:00-12:50

Instructor Information

- Title & Name of Instructor: Dr. Joann J Ordille, Associate Professor of Professional Practice
- Office Hours:
    - *o* Livingston Campus Office, BRR-5066
        - Tuesday, 5:30 pm – 6:30 pm
        - Friday, 2:30 pm – 4:30 pm
    - *o* Newark Campus Office, 1WP-1015A
        - Wednesday, 1:00 pm – 2:00 pm
    - *o* Students can also attend these office hours via Zoom using the office hour link available on the Canvas home page.
- Email: jo531@scarletmail.rutgers.edu
- Office Phone Number:
    - o 848-445-3243 on Livingston Campus
    - o Use my Livingston office phone number to contact me during my Newark office hours.

Experiential Learning

This course integrates experiential learning to connect academic concepts with real-world business applications. Learning methods deployed in this course include case-based analysis and project-based work. These approaches enable students to apply theories to realistic scenarios, develop practical skills, and address complex business challenges.

Course Description

This course focuses on research and applications for storing, managing and processing big data in the cloud. The course addresses each stage in the Data Science/Machine Learning (ML) Data Pipeline. It covers a variety of database architectures for big data: data warehouses, data lakes and data lakehouses. It illustrates these architectures using a variety of popular cloud providers.

This course provides an opportunity to learn about the end-to-end process of acquiring, preparing, storing, processing and using big data in data science. Students apply their learning on popular cloud platforms. The course topics include how to address the five V's of Big Data: volume, variety, velocity, veracity, and value. We also address how to maintain the virtue of our data, a sixth V if you will, by addressing issues of security, privacy, and social responsibility.

Advanced database research has produced a collection of powerful and successful NoSQL (Not Only SQL) database systems and data processing techniques to address the challenges presented by big data. This course covers key-value stores, wide-column databases, document databases, graph databases and streaming data systems.

Key-value stores form the foundation for fast, incrementally scalable, distributed processing of Internet shopping carts, user information, and product information. We discuss Amazon's DynamoDB as an example of key-value stores. Wide-column databases support fast information storage and retrieval for massive sparse matrix applications such as search engines, personalization of services, analytics, and email. Google's BigTable and Facebook's Cassandra are our examples of wide-column databases. Our example of a document database is MongoDB. MongoDB undergirds the high performance of many web applications. It is currently the most popular NoSQL database. Graph databases support analyzing social media relationships, transportation systems, and disease outbreaks. These databases increasingly find a role in automating machine learning pipelines, and are illustrated by Neo4j and Pregel. Data generated at high velocity such as data generated by sensors in the Internet of Things (IOT) require a streaming data system. We dive into these systems using Spark Streaming, Google Dataflow, Apache Beam, and Amazon Kinesis.

We examine how these databases conform to the CAP Theorem by making tradeoffs between data consistency, availability, and resilience to network partitioning in order to achieve scale. We also explore how underlying technologies like MapReduce and Spark make these systems possible.

During Fall 2024, free access to Amazon Web Services (AWS), the Amazon Cloud Platform, is provided to students in this course as part of the AWS Academy Program. Free access to MongoDB is provided through MongoDB Atlas, and to Neo4j through Neo4j AuraDB.

In this course, class meetings will be a combination of lecture, discussion, team presentations and group exercises. Students will build and manipulate various databases and cloud services. Students will build applications on their own laptops and in the cloud.

Course Delivery Mode

All sections of this class are in-person.

Course Materials

- Required Textbooks:
    - Akidau, T., Chernyak, S., & Lax, R. (2018). *Streaming systems: the what, where, when, and how of large-scale data processing*. O'Reilly Media, Inc. Available from the O'Reilly Database through the Rutgers Library.
    - Carpenter, J. & Hewitt, E. (2022). *Cassandra: the definitive guide* (Revised 3rd ed.). O'Reilly Media, Inc. Available from the O'Reilly Database through the Rutgers Library.
    - Harrison, G. (2016). *Next generation databases: NoSQL, newSQL, and big data*. Apres. Available from the O'Reilly Database through the Rutgers Library.
    - Perkins, L., Redmond, E., & Wilson, J. (2018). *Seven databases in seven weeks: a guide to modern databases and the NoSQL movement (2nd ed.)*. Pragmatic Bookshelf. Available from the O'Reilly Database through the Rutgers Library.
- Recommended books:
    - Damji, J., Lee, D., Wenig, B., & Das, T. (2020). *Learning Spark: lightning-fast big data analytics* (2nd ed.) O'Reilly Media, Inc. Available from the O'Reilly Database through the Rutgers Library.
    - Lin, J., & Dyer, C. (2010). Data-intensive text processing with MapReduce. *Synthesis Lectures on Human Language Technologies*, *3*(1), 1-177.
- Articles in conferences proceedings, journals and professional publications are used in this course as described in the schedule below.
- Check Canvas at Rutgers and your Rutgers Scarlet Mail email account regularly for course updates and announcements.

Learning Goals and Objectives

For Master of Information Technology and Analytics (MITA) Program, this course satisfies the following Rutgers Business School goals and objectives:

- Business technology knowledge. Students graduating with a Master of Information Technology degree will be able to demonstrate business technology knowledge. Students will demonstrate:
    - Understanding of the current practices and technology used in businesses.
    - Ability to analyze and solve complex business problems with cutting edge technology.
- Information technology knowledge. Students graduating with a Master of Information Technology degree will be able to demonstrate information technology knowledge. Students will demonstrate:
    - Understanding of basic information technology concepts.
    - Ability to analyze and solve information technology problems.

- Critical thinking skills. Students graduating with a Master of Information Technology degree will be able to understand complex business situations and provide solutions to improve current business practices. Students will demonstrate:
    - Ability to identify problems in a situation.
    - Ability to find innovative solutions.
- Communication skills. Students graduating with a Master of Information Technology degree will be able to effectively communicate in a way that demonstrates sensitivity to an audience's needs. Students will demonstrate:
    - Ability to communicate information in a clear concise manner.
    - Ability to communicate relatively complex ideas in an understandable manner.

Prerequisites

Students taking this course should have knowledge of relational database systems, including database normalization, entity relationship diagrams and design, and SQL, and experience in computer programming.

Tentative Course Schedule*

Date	Week	Topic	Readings, Quizzes and Due Dates
9/3	1	Introduction to Course and Cloud	An excerpt from Lisdorf, A. (2021). "Introduction" in *Cloud Computing Basics: A Non-Technical Introduction.* Apres, pp. xiii-xv (3 pages). How Cloud Computing Became a Big Tech Battleground. (2019). Wall Street Journal. (4 minutes, 16 seconds). Mell, P., & Grance, T. (2011). Section 2 in The NIST definition of cloud computing. National Institute of Standards, Publication 800-145, pp. 2-3. (2 pages). Ranger, S. What is cloud computing? Everything you need to know about cloud explained. (2022). ZDNet. (14 pages). How Cloud Giants Amazon, Google, and Microsoft Got Even Bigger. (2022). Wall Street Journal Tech News Briefing. (8 minutes, 6 seconds) Cloud Computing Isn't as Cost Effective as Hoped. So What's Next? (2022). Wall Street Journal Tech News Briefing. (6 minutes, 17 seconds) Laberis, B. (2019). The disruptive force of cloud native. Natunix. (4 pages). Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., ... & Zaharia, M. (2010). A view of cloud computing. *Communications of the ACM, 53*(4), 50-58. (9 pages).

Date	Week	Topic	Readings, Quizzes and Due Dates
9/10	2	Cloud Architectures and Software Processes. Putting it together with AWS.	**Put what we covered last time into practice:** Introduction, AWS Academy Cloud Foundations Module 1: Cloud Concepts Overview, Module 2: Cloud Economics and Billing, Module 3: AWS Global Infrastructure Overview, and Module 4: AWS Cloud Security including Lab 1 and Knowledge Checks. **First preliminaries for database assignments:** See Canvas for an assignment to install tools. **Preparing for today's class:** For IBM Cloud resources, feel free to skip IBM-specific product information. IBM Cloud Team (2021). Containers vs. virtual machines (VMs): What's the difference? IBM. (4 pages plus 13 minutes and 19 seconds of video). IBM Cloud Education (2021). Docker. IBM. (12 pages plus 10 minutes and 59 seconds of video). IBM Cloud Education (2020). Continuous Integration. (8 pages plus 6 minutes and 20 seconds of video). IBM Cloud Education (2019). Continuous Deployment. (7 pages plus 7 minutes and 36 seconds of video). Savor, T., Douglas, M., Gentili, M., Williams, L., Beck, K., & Stumm, M. (2016, May). Continuous deployment at Facebook and OANDA. In *2016 IEEE/ACM 38th International Conference on Software Engineering Companion (ICSE-C)* (pp. 21-30). IEEE. (10 pages) Watch the video of Tony Savor presenting the paper at the Canadian Tech at Scale Conference (29 minutes and 26 seconds). **Recommended Reading and Video on Orchestration:** IBM Cloud Education (2019). What is Kubernetes? (14 pages plus 11 minutes and 57 seconds of new video. One of the videos is also in the Docker reference above.).

Date	Week	Topic	Readings, Quizzes and Due Dates
9/17	3	Big Data, Data Warehouses, Data Lakes, and Data Pipelines	**Put what we covered last time into practice:** AWS Academy Cloud Foundations Module 5: Networking and Content Delivery and Module 6: Compute including Labs 2 and 3, Activities, and Knowledge Checks. **Second preliminaries for database assignments:** See Canvas for an assignment on the Unix/Linux shell. **Preparing for today's class:** Ellingwood, J. (2016). An Introduction to Big Data Concepts and Terminology. DigitalOcean. (6 pages) Dageville, B., Cruanes, T., Zukowski, M., Antonov, V., Avanes, A., Bock, J., et al. (2016, June). The snowflake elastic data warehouse. In *Proceedings of the 2016 International Conference on Management of Data* (pp. 215-226). (12 pages) Armenatzoglou, N., Basu, S., Bhanoori, N., Cai, M., Chainani, N., Chinta, K., et al. (2022, June). Amazon Redshift re-invented. In *Proceedings of the 2022 International Conference on Management of Data* (pp. 2205-2217). (13 pages) Nargesian, F., Zhu, E., Miller, R. J., Pu, K. Q., & Arocena, P. C. (2019). Data lake management: challenges and opportunities. *Proceedings of the VLDB Endowment*, *12*(12), 1986-1989. (4 pages) Armbrust, M., Ghodsi, A., Xin, R., & Zaharia, M. (2021, January). Lakehouse: a new generation of open platforms that unify data warehousing and advanced analytics. In *Proceedings of CIDR* (Vol. 8). (7 pages) **Recommended Reading on Data Cubes:** Han J, Kamber M, Pei J. (2012). Sections through 4.4 from "Chapter 4: Data Warehousing and Online Analytical Processing" in *Data Mining Concepts and Techniques* (3rd ed.), pp. 125-165. Elsevier. (41 pages)

Date	Week	Topic	Readings, Quizzes and Due Dates
9/24	4	Big Data Case Studies	**Put what we covered last time into practice:** AWS Academy Cloud Foundations Module 7: Storage including Lab 4 with Knowledge Check, and AWS Academy Data Engineering Module1: Welcome to AWS Academy Data Engineering, Module 2: Data-Driven Organization, Module 3: The Elements of Data, and Module 8: Storing and Organizing Data including the lab in Module 2 and the lab in Module 8 with Knowledge Checks. Module 1 is missing its video, so just flip through the slides. Each student team presents a case study of big data use in a company from Marr, B. (2016). *Big data in practice: how 45 successful companies used big data analytics to deliver extraordinary results*. John Wiley & Sons. Available from the O'Reilly Database through the Rutgers Library.
10/1	5	**Quiz 1**, Big Data Processing, MapReduce, Spark	**Preparing for today's class:** Harrison, G. (2016). Chapter 2: Google, Big Data, and Hadoop. Published in *Next generation databases: NoSQL, newSQL, and big data*, pp. 21-37. Apres. (17 pages) Available from the O'Reilly Database through the Rutgers Library. Dean, J., & Ghemawat, S. (2008). MapReduce: simplified data processing on large clusters. *Communications of the ACM*, *51*(1), 107-113. (7 pages) Zaharia, M., Xin, R. S., Wendell, P., Das, T., Armbrust, M., Dave, A., ... & Stoica, I. (2016). Apache spark: a unified engine for big data processing. *Communications of the ACM*, *59*(11), 56-65. (10 pages) Spark on Google Colab (single node) Spark on AWS EMR Cluster (multiple node cluster) **Recommended Reading on MapReduce Programming:** Lin, J., & Dyer, C. (2010). Chapter 1: MapReduce basics. Published in Data-intensive text processing with MapReduce. *Synthesis Lectures on Human Language Technologies*, *3*(1), 18-38

Date	Week	Topic	Readings, Quizzes and Due Dates
10/8	6	Hive, Distributed Databases, CAP, Scalability and Elasticity	**Put what we covered last time into practice:** See Canvas for MapReduce and Spark assignments. AWS Academy Data Engineering Module 9: Processing Big Data with two labs and Knowledge Checks. **Preparing for today's class:** Garcia-Molina, H., Ullman, J., & Widom, J. (2009). 20.1 Parallel Algorithms on Relations. Published in *Database Systems: The Complete Book* (2nd ed.), pp. 985-993, 1008-1013. Pearson Education. (8 pages) Thusoo, A., Sarma, J. S., Jain, N., Shao, Z., Chakka, P., Zhang, N., ... & Murthy, R. (2010, March). Hive-a petabyte scale data warehouse using Hadoop. In *2010 IEEE 26th international conference on data engineering (ICDE 2010)* (pp. 996-1005). IEEE. (10 pages) Garcia-Molina, H., Ullman, J., & Widom, J. (2009). 20.3 Distributed Databases, 20.3.1 Distribution of Data, 2.3.2 Distributed Transactions, 2.3.3 Replication, 20.5 Distributed Commit (including subsections 20.5.1, 20.5.2, and 20.5.3). Published in *Database Systems: The Complete Book* (2nd ed.), pp. 997-999, 1008-1013. Pearson Education. (9 pages) Carpenter, J. & Hewitt, E. (2022). Beyond relational databases. Published in *Cassandra: the definitive guide* (Revised 3rd ed.), 1-16. O'Reilly Media, Inc. (16 pages) Available from the O'Reilly Database through the Rutgers Library. Abadi D. (2012). Consistency Tradeoffs in Modern Distributed Database System Design: CAP is Only Part of the Story. Computer (Long Beach, Calif). 45(2):37-42. doi:10.1109/MC.2012.33. (6 pages) **Recommended Reading on MapReduce Programming Models:** Olston, C., Reed, B., Srivastava, U., Kumar, R., & Tomkins, A. (2008, June). Pig latin: a not-so-foreign language for data processing. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data* (pp. 1099-1110).

Date	Week	Topic	Readings, Quizzes and Due Dates
10/15	7	Intro to Key-Value Databases with Amazon's Dynamo. Intro to Wide-Column Databases with Google's BigTable	**Put what we covered last time into practice:** Data Engineering Modules 4: Design Principles and Patterns for Data Pipelines and Module 5: Securing and Scaling the Data Pipeline including the lab in Module 4, and Knowledge Checks. Note that the "Cloud security review" recording in Module 5 overlaps with Module 4 in the AWS Academy Cloud Foundations Course. **Preparing for today's class:** Harrison, G. (2016). Chapter 3: Sharding, Amazon and the Birth of NoSQL. Published in *Next generation databases: NoSQL, newSQL, and big data*, pp. 39-51. Apres. (13 pages) Available from the O'Reilly Database through the Rutgers Library. DeCandia, G., Hastorun, D., Jampani, M., Kakulapati, G., Lakshman, A., Pilchin, A., ... & Vogels, W. (2007). Dynamo: Amazon's highly available key-value store. Published in the Proceedings of the 2007 Symposium on Operating Systems (SOSP '07), *ACM SIGOPS operating systems review*, *41*(6), 205-220. (16 pages) Krzyzanowski, P. (2021). BigTable: A NoSQL wide-column single-table database. (8 pages) Chang, F., Dean, J., Ghemawat, S., Hsieh, W. C., Wallach, D. A., Burrows, M., ... & Gruber, R. E. (2008). Bigtable: A distributed storage system for structured data. *ACM Transactions on Computer Systems (TOCS)*, *26*(2), 1-26. (27 pages) **Recommended Readings. The second article is from Google on building a relational-style (NewSQL) database called Megastore on top of BigTable. Megastore powers Google's App Engine. If you skip Section 4 through 4.9, you can still get the gist. If you want to read Section 4, best to read the article about Paxos first.** Krzyzanowski, P. (2018). Understanding Paxos: Asynchronous Fault-Tolerant Consensus. (9 pages) Baker, Jason, Chris Bond, James C. Corbett, J. J. Furman, Andrey Khorlin, James Larson, Jean-Michel Leon, Yawei Li, Alexander Lloyd, and Vadim Yushprakh. (2011). Megastore: Providing scalable, highly available storage for interactive services. Published in the *Proceedings of the 5th Biennial Conference on Innovative Data Systems Research (CIDR '11)*, 223-234. (12 pages)

Date	Week	Topic	Readings, Quizzes and Due Dates
10/22	8	Cassandra, introduced by Facebook in 2007, combining Wide-Column and Key-Value Database Features. Extended Microservice Example using Cassandra.	**Put what we been covering into practice:** AWS Academy Cloud Foundations Module 8: Databases with Knowledge Check and Lab 5. **Preparing for today's class:** Carpenter, J. & Hewitt, E. (2022). Chapter 2: Introducing Cassandra. Published in *Cassandra: the definitive guide* (Revised 3rd ed.), 17-31. O'Reilly Media, Inc. (15 pages) Available from the O'Reilly Database through the Rutgers Library. Carpenter, J. & Hewitt, E. (2022). Chapter 4: The Cassandra Query Language and Chapter 5: Data Modeling in Introducing Cassandra. Published in *Cassandra: the definitive guide* (Revised 3rd ed.), 55-106. O'Reilly Media, Inc. (52 pages) Available from the O'Reilly Database through the Rutgers Library.
10/29	9	Quiz 2, Document Stores and MongoDB	**Preparing for today's class:** Harrison, G. (2016). Chapter 4: Document databases. Published in *Next generation databases: NoSQL, newSQL, and big data*, pp. 53-63. Apres. (11 pages) Available from the O'Reilly Database through the Rutgers Library. Harrison, G. (2016). Chapter 8: Distributed database patterns. Published in *Next generation databases: NoSQL, newSQL, and big data*, pp. 110-115. Apres. (5 pages) Read the subsection on MongoDB Sharding and Replication only. Available from the O'Reilly Database through the Rutgers Library. Harrison, G. (2016). Chapter 11: Languages and programming interfaces. Published in *Next generation databases: NoSQL, newSQL, and big data*. pp. 173-175. Apres. (3 pages) Read the subsection on MongoDB only. Available from the O'Reilly Database through the Rutgers Library. Copeland, R. (2013). To Embed or Reference. Published in MongoDB Applied Design Patterns: Practical Use Cases with the Leading NoSQL Database, pp. 3-14. O'Reilly Media, Inc. (12 Pages) Available from the O'Reilly Database through the Rutgers Library. Note: MongoDB added transactions in Version 4.0 (2018) with enhancements in Version 4.2 (2019). Schultz, W., Avitabile, T., & Cabral, A. (2019). Tunable consistency in mongodb. *Proceedings of the VLDB Endowment*, *12*(12), 2071-2081. (11 pages)

Date	Week	Topic	Readings, Quizzes and Due Dates
11/5	10	Graph Databases	**Put what we've been covering into practice:** AWS Academy Cloud Foundations Module 9: Cloud Architecture and Module 10: Autoscaling and Monitoring with Knowledge Checks and Lab 6. Complete the course assessment and feedback. You will receive a badge from AWS Academy. Congratulations, you have completed the AWS Academy Cloud Foundations Course. See Canvas for MongoDB Assignment due next week. The exercise is based on the following: Perkins, L., Redmond, E., & Wilson, J. (2018). Chapter 4: MongoDB. Published in *Seven databases in seven weeks: a guide to modern databases and the NoSQL movement*, pp. 93-133. Pragmatic Bookshelf. (40 pages) Available from the O'Reilly Database through the Rutgers Library. **Preparing for today's class:** Harrison, G. (2016). Chapter 5: Tables are not your friends: Graph databases. Published in *Next generation databases: NoSQL, newSQL, and big data*., pp.65-74. Apres. (10 pages) Available from the O'Reilly Database through the Rutgers Library. Malewicz, G., Austern, M. H., Bik, A. J., Dehnert, J. C., Horn, I., Leiser, N., & Czajkowski, G. (2010, June). Pregel: a system for large-scale graph processing. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*, 135-146. (12 pages) Check out the Stanford Network Analysis Project (SNAP) for some ideas about what can be represented in graphs and the results that can be obtained by analyzing them. **Recommended Reading on the Bulk Synchronous Parallel (BSP) Model:** Valiant, L. G. (1990). A bridging model for parallel computation. *Communications of the ACM*, *33*(8), 103-111.

Date	Week	Topic	Readings, Quizzes and Due Dates
11/12	11	Sources of Velocity. Streaming Systems.	**Put what we've been covering into practice:** MongoDB Assignment due this week. See Canvas for Neo4j Assignment due the week after Thanksgiving. Perkins, L., Redmond, E., & Wilson, J. (2018). Chapter 6: Neo4J. Published in *Seven databases in seven weeks: a guide to modern databases and the NoSQL movement*, pp. 177-209. Pragmatic Bookshelf. (33 pages) **Preparing for today's class:** Kleppmann, M. (2016). Chapter 1. Events and stream processing. Published in *Making sense of stream processing*, 1-37. (38 pages) O'Reilly Media, Inc. Akidau, T., Bradshaw, R., Chambers, C., Chernyak, S., Fernández-Moctezuma, R. J., Lax, R., ... & Whittle, S. (2015). The dataflow model: a practical approach to balancing correctness, latency, and cost in massive-scale, unbounded, out-of-order data processing. Published in *Proceedings of the VLDB Endowment* (Vol. 8), 1792-1803. (12 pages) Akidau, T., Chernyak, S., & Lax, R. (2018). "Chapter 2: The What, Where, When, and How of Data Processing" in *Streaming systems: the what, where, when, and how of large-scale data processing*, pp. 25-57. (33 pages) O'Reilly Media, Inc. Available from the O'Reilly Database through the Rutgers Library. Read it with the online animated figures.

Date	Week	Topic	Readings, Quizzes and Due Dates
11/19	12	Veracity and Virtue, Data Ingestion and Preparation in the Pipeline	**Preparing for today's class:** Stoyanovich, J., Howe, B., Abiteboul, S., Miklau, G., Sahuguet, A., & Weikum, G. (2017, June). Fides: Towards a platform for responsible data science. In *Proceedings of the 29th International Conference on Scientific and Statistical Database Management*, 1-6. (6 pages) Available from the Rutgers Library: https://dl-acm-org.proxy.libraries.rutgers.edu/doi/abs/10.1145/3085504.3085530 Werder, K., Ramesh, B., & Zhang, R. (2022). Establishing data provenance for responsible artificial intelligence systems. *ACM Transactions on Management Information Systems (TMIS)*, *13*(2), 1-23. (23 pages) Available from the Rutgers Library: https://bit.ly/3R5JH67 AWS Academy Data Engineering Module 6: Ingesting and Preparing Data and Module 7: Ingesting by Batch or by Stream with two labs as you encounter them and Knowledge Checks.
11/26	13	No Class. Change in Designation Day to Friday classes.	Happy Thanksgiving!
12/3	14	**Quiz 3.** Final Quiz and Project Q and A	

Date	Week	Topic	Readings, Quizzes and Due Dates
12/10	15	Early Short Project Presentations. Data Pipelines and ML, Data Analysis and Visualization. Automated Pipelines	This is also the last day of classes. **Note: Depending on the number of early presentations, the topics for this class meeting may change.** **Preparing for today's class:** AWS Academy Data Engineering Modules 10: Processing Data for ML, Module 11: Analyzing and Visualizing Data, and Module 12: Automating the Pipeline including the Module 11 Lab and Module 12 Lab, and Knowledge Checks. Complete the course assessment. You will receive a badge from AWS Academy. The Data Engineering Course is partial preparation for advanced AWS certifications. Check out the slides for Module 13 for more information. Leskovec, J. (2023, June). Databases as Graphs: Predictive Queries for Declarative Machine Learning. In *Proceedings of the 42nd ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems* (pp. 1-1). (1 page) Leskovec, J. (2023). Graphs, Databases and Machine Learning. Innovation Award Presentation at the *29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. (26 minutes 53 seconds)

***Tentative schedule, subject to change. Later topics may be abbreviated if time does not allow full treatment. Check Canvas for the most up to date information on the schedule, readings and assignments.**

Important Dates

Important dates for the Fall 2025 Semester from the RBS Academic Calendar:

- Start of Classes: Tuesday, September 2, 2025
- End of Drop/Add Period: Thursday, September 11, 2025
- Last Day to Withdraw with a W Grade: Monday, October 27, 2025
- Thanksgiving Break: Thursday, November 27 – Sunday, November 30, 2025
- End of Classes: Wednesday, December 10, 2025
- Exam Schedule: See the Exam Dates and Policy Section.

Course Expectations

Feedback and Response Expectations

- If you have a question that you think is shared by your classmates, please ask the question by posting a discussion topic on Canvas. I will answer your question there, so everyone will learn the answer. You can download a Canvas app for your cell phone, so you can participate in discussions from anywhere.

- An email to my Scarlet Mail address (jo531@scarletmail.rutgers.edu) is the best way to reach me about our course. Unlike Canvas Direct Messages, Scarlet Mail provides context for continuing conversations. Emailing joann.ordille@rutgers.edu might result in a response delay, because it does not notify my cell phone. Scarlet Mail does notify my cell phone.
- Set up Canvas to notify you of course announcements. Check Canvas often. You are responsible for meeting the deadlines and fulfilling the requirements posted on Canvas.
- I will email you at your Scarlet Mail address. If you do not have a Scarlet Mail address, visit NetID Management and Service Activation to set one up. You can also use this website to have your Scarlet Mail forwarded to your preferred email address.
- Email/Direct Messaging Response Times: I will do my best to respond to your emails quickly, often within hours, but at a maximum within 24 hours weekdays and 48 hours on weekends. During Thanksgiving Break, email responses may be delayed until class resumes. Please remind me if you do not hear back from me within this time.
- Graded Materials Return Times: Most of your quiz questions will be graded automatically by Canvas. Since I have several large classes, I will strive to have your other assignments and tests graded within two weeks but three weeks may be required.
- In this course you will be expected to complete several tasks including:
    - doing online reading and assignments on time
    - downloading and uploading documents and code using Canvas
    - accessing online resources including tools, articles, tutorials, and videos
    - creating videos
    - communicating via Canvas Discussions
    - completing synchronous quizzes via Canvas using the Respondus Lockdown Browser
    - participating in discussions and activities in in-person class meetings

Attendance and Preparation

- Learning is a joint responsibility of teacher and student. Both teachers and students are expected to prepare for and attend class, and actively participate in the learning process. If I will miss class, my department chair or I will send you notice via email and Canvas as far in advance as possible. During inclement weather, always check Canvas announcements before travelling to class.
- If you need to miss class, please report your absence in advance using the Student Self-Reporting Absence System (SSRA). If your absence is due to religious observance, a Rutgers-approved activity, illness, or family emergency/death and you seek makeup work, also send me an email with full details and supporting documentation within two days of your first absence. If you miss an exam or quiz without first negotiating other arrangements, you will NOT be entitled to a make-up and will receive a zero UNLESS you can provide documentation of illness or a family emergency. Excused absences will not be counted toward your attendance grade.
- Expect me to arrive on time for each class session. I expect the same of you.
- Expect me to remain for the entirety of each class session. I expect the same of you.

- Expect me to prepare properly for each class session. I expect the same of you. Complete all background reading and assignments. You cannot learn if you are not prepared. The minimum expectation is that for each 3 hour class session, you have prepared by studying for at least twice as many hours.
- Actively engage in class learning by asking or answering questions, expressing an opinion about a topic of discussion, meeting with me during office hours, interacting with me or other class members through Canvas, participating in your team and reporting on activities in a project or other activity. Stay focused and involved. You cannot learn if you are not paying attention.
- Reflect on your learning process. What do you understand? What is unclear? Seek clarity and understanding. Consult class resources, reach out in class, through email, or during office hours to find answers to your questions and build your knowledge and skills to achieve success.

Classroom Conduct

Research has shown that students learn better in a community with their peers. I hope to help you form that community by creating teams. These teams will participate in class in group activities. They will collaborate in reading and discussing research papers in preparation for class meetings. Teams will submit summaries of their discussions or be required to ask or answer questions in class. Each team will also have the responsibility for presenting a case study during one of the class meetings. Teams will consult with me in advance of their presentation, and every member must take an active role in doing the presentation.

In class, we will sometimes have active discussion sessions. A series of students may be called upon (cold called) to answer questions or contribute an opinion. If what you know is insufficient to answer, you are permitted to pass.

Course Policies

Exam Dates and Policies

There are three quizzes in this course. The quizzes are closed book and in-person. The quizzes focus on the most recent material, but the material builds through the term and is in that sense cumulative. The quizzes occur approximately every four weeks.

During quizzes, the following rules apply:

- In-person quizzes are closed book with restricted online access enforced by the Respondus Lockdown Browser.
- If you qualify for any accommodation that will influence testing procedures, have the Office of Disability Services send me a letter at the start of the semester. Reach out to discuss the accommodation with me.
- All tests are in-person.
- If you take all or part of an in-person test while not present in class in person, you will be reported for a violation of academic integrity and receive a zero for the test.

- If you leave the classroom without submitting your test, it will be considered an attempt to complete the test online while not present in class. You will be reported for a violation of academic integrity and receive a zero for the test.
- There are no situations that qualify for taking a test while not present in-person. If you cannot take the test in-person at the scheduled time, an in-person makeup test may be granted in accordance with policy explained in the Attendance and Preparation Section.
- If you are granted a makeup quiz, the quiz may have a different format from the one given to the class at the discretion of the instructor. The quiz might, for example, be an oral test.
- The Rutgers Business School Dean's Office has informed the faculty that incompletes are only to be given for documented cases that prevent you from completing requirements of a course. The first step in producing such documentation is to contact the Dean of Students Office and ask for help. See the Dean of Students Web Page for more information on requesting help.
- If you miss the final quiz, project presentation or project submission, and do not have permission to receive an incomplete from the Dean of Students, you will be given a zero, and a final grade will be assigned in lieu of an incomplete. If a makeup is granted, you will need to complete the in-person components of the makeup at a time convenient to your instructor, typically at the start of the next term. After the makeup is completed and graded, your grade in the course will be updated.
- In the instructions for each quiz, the quiz will explicitly list the permitted material and technologies that you can use when taking the quiz. If you access any material or technology that is not permitted, you will be reported for a violation of academic integrity and receive a zero for the test.
- You can only use authorized technologies during a quiz. For example, smart glasses are not permitted. If you need prescription lenses, be sure to obtain a classic pair of glasses that does not include any type of smart technology. If you use any unauthorized technologies, such as cell phones, smart glasses or earpieces, during the quiz, you will be reported for a violation of academic integrity and receive a zero for the quiz.
- Cell phones must be turned off and placed in front of you face down during the quiz.
- Use the bathroom prior to the test start time.
- Your quiz will receive a grade of zero if you do not sign the Honor Pledge.

Grading Policy

Course grades, including final grades, will be posted on Canvas. Grades are determined based on the following categories of work:

- **Class Attendance.** Attendance will be taken with Qwickly. If you attend less than 75% of the class meeting on a particular day, you will not receive credit for attending on that day. Your attendance grade will be the percentage of class meetings you attend. Excused absences will not be counted toward your grade. Attendance is worth 2% of your grade.
- **Team Participation:** As described in the Classroom Conduct Section, you will be assigned to a team to learn collaboratively with your peers. Your contribution to your team counts for 2% of your grade.

- **Team Class Presentation:** As described in the Classroom Conduct Section, each team will also have the responsibility for presenting a case study during one of the class meetings. Teams will consult with me in advance of their presentation, and every member must take an active role in doing the presentation. This presentation is worth 6% of your grade.
- **Homework:** "Put it into practice" activities described in the timetable may have deliverables, and other exercises will be assigned as needed. This category is worth 30% of your final grade.
- **Individual Project:** You are required to do an individual term project. The project is worth 30% of your grade.

    - **Survey paper.** (Read at least 6 papers on the topic.)

      Use Google Scholar, ACM Portal and DBLP to find papers, focusing on those published in the following conferences: VLDB, SIGMOD, and ICDE. Depending on your topic, other conferences such as SOSP or CIDR may also be appropriate. Feel free to see me for guidance on conference selection.
      Write a survey that includes an introduction, problem definition (including motivation and application domain), summary of techniques developed in each paper to address the problem, global view of the papers covered, and future work suggestions. The length should be limited to and not exceed 6 pages in ACM conference format (references can be on additional pages).
      You will present your work, and it will be evaluated on (a) understanding of the topic, (b) presentation and structure, and (c) critique of the research covered.
    - **Own research.**
      Proceed in the same manner as for the survey option above. In addition, identify a new research problem in the area and develop your own solution. Submit a paper describing your work. Your paper should include a motivation that shows how your work addresses a problem that related work did not address. It should compare your solution with related work. If your work includes experimental results, be sure to make a clear separation between the presentation of the measurements and your interpretation of them. You will present your work. Your work will be evaluated for originality and novelty, and convincing argument or experimental results. In this case, the comprehensiveness of the survey becomes secondary.
    - **Build a prototype.**
      Identify a research problem and examine existing solutions from papers in the literature. Implement one of the solutions, as found in a rank one conference (i.e., VLDB, SIGMOD, ICDE, SOSP) or premium journal paper (i.e., ACM TODS, VLDB Journal, IEEE TKDE, ACM TOCS). Feel free to see me for guidance on conference/paper selection. Write a 4-6 pages report (references can be on additional pages) using ACM format as above. Include a discussion of the problem and the solution, and your experimental results. Try to reproduce some of the results in the paper. You may use artifacts provided by the authors of the

paper as part of your evaluation. Submit the report along with a zip file of your code and any code used from the authors of the paper. Include instructions for installing and running your system/solution in a file called README in your zip submission. Your report should explain whether you confirmed the published results or found some discrepancy, and what your results mean. You will present and demonstrate your prototype, and the work will be evaluated on (a) report quality and (b) demonstration effectiveness.

- **Master's Students Only: Build an application.**

  Identify an application of the database systems, data processing techniques or the cloud related to the course content. Build an application of the system/techniques. Write a 4-6 pages report (references can be on additional pages) using ACM format as above. Include a discussion of the problem your application solves and the solution. Discuss how your work illustrates, extends or diverges from the research in the area discussed in the course. Discuss what you learned and your suggestions for future work. Building and evaluating an application using two different systems/techniques from the course, and then comparing the results and your experiences can lead to a very compelling report. Submit the report along with a zip file of your code. Include instructions for installing and running your system/application in a file called README in your zip submission. You will be called to demonstrate your application, and the work will be evaluated on (a) report quality and (b) demonstration effectiveness.

- **Your project must be approved.** To obtain approval, submit a proposal for your project by the deadline given in class.

- **What if I'm late completing the Individual Project?** If you are unprepared to discuss or demonstrate your work during the designated time at the end of term, you will lose the points for that part of the project grade. For the remainder, late submission of your work will be penalized as follows:
  - 1 day late, grace period with no points off
  - 2-3 days late, 3% off per day
  - 4th day late, 4% off
  - 5-10 days late, 5% off per day
  - 11 or more, 10% off per day until no points are available and the grade is zero.

- **Quizzes:** There are three in person quizzes. They are closed-book and worth 30% of your grade.

The following summarizes how each category of work contributes to your final numerical grade:

Category of Work	Percentage of Grade
Class Attendance	2%
Team Participation	2%

Category of Work	Percentage of Grade
Team Class Presentation	6%
Homework	30%
Quizzes	30%
Individual Project	30%

Grades will be assigned as follows from your final numeric grade for students in the master's degree section of the course (22:544:645:01):

- A: 93-100
- A-: 90-92
- B+: 87-89
- B: 83-86
- B-: 80-82
- C+: 77-79
- C: 73-76
- C-: 70-72
- D: 60-69
- F: 0-59

Other important notes:

- In addition to the ability to answer homework type problems, quizzes will also test your conceptual understanding of material, your ability to integrate what you have learned and analyze it, and ability to apply what you learned and extend it. Are you able to discuss the tradeoffs between different strategies for addressing database consistency? Are you able to suggest new approaches to solving database problems? Can you advise a company on big data strategies?
- There is NO extra credit. Plan to earn enough points to pass the course.
- Grades will be posted on Canvas. Time for returning grades is discussed in the Feedback and Response Expectations Section.
- Grades are not subject to negotiation. If you feel we made an error, submit your written argument to me within one week of receiving your grade, including your final grade. Clarify the precise error made and provide supporting documentation. If we made an error in grading, I will gladly correct it.

Artificial Intelligence Use

Use of AI such as ChatGPT is only permitted to help you brainstorm ideas and see examples. All material you submit for assignments must be your own. Use of AI such as ChatGPT is not permitted during quizzes or exams.

Be aware that AI software, such as ChatGPT, may give a false answer to an assignment, or a truthful statement that does not satisfy the requirements of the assignment. Using either of these types of answers would lose full or partial credit on the assignment. You are responsible for the correctness and appropriateness of your answers, not ChatGPT.

Handing in assignments that are generated by AI is a violation of this policy and of the academic integrity policy. Since you are responsible for completely understanding and creating any assignment that you submit and not just handing in something you find somewhere or generate with AI, you may be tested on the content of your submission.

Academic Integrity

I do **not** tolerate cheating or academic dishonesty of any kind. Students are responsible for understanding and adhering to the Rutgers Academic Integrity Policy. I will strongly enforce this policy and pursue *all* violations. On all examinations and assignments, students must sign the RU Honor Pledge, which states, "On my honor, I have neither received nor given any unauthorized assistance on this examination or assignment." Failure to sign the honor statement will result in a zero for the quiz, examination or assignment. I will screen all written assignments through Turnitin plagiarism detection services that compare the work against a large database of past work. Don't let cheating or plagiarism destroy your hard-earned opportunity to learn and advance. See the RBS Academic Integrity Resource Page for more details.

Student Code of Professional Conduct

Rutgers Business School is recognized for its high-quality education. Maintaining the caliber of classroom excellence, whether in person or online, requires students to adhere to the same behaviors that are expected in professional career environments. These include the principles outlined in the RBS Student Code of Professional Conduct.

Support Services

Technology Support and Information

- Learning Management System: Canvas at Rutgers University
- Hardware and software requirements*:*
    - a webcam for recording an introductory video,
    - a laptop in class for pre-announced working sessions or taking quizzes,
    - at least 75 MB on Windows or 120 MB on Macs to install the Respondus Lockdown Browser, and
    - 0.5 GB free space on a Windows laptop or Mac to install tools, create programs, and write the project paper.
- Below are the minimum hardware requirements recommended by OTIS. These specs will allow student systems to capably support a full Windows or macOS environment with Office 365, RBS course-specific applications, and virtual computing environments:
    - Intel® Core™ i5 processor (10th generation or newer) or Apple M1 (or newer) processor
    - Windows 11 Professional, or macOS 11 (Big Sur) or newer
    - 8 GB of RAM (16 GB recommended)
    - 256 GB solid-state drive (SSD) or larger
    - 720p HD webcam (1080p recommended)
    - Internal microphone

    - Reliable internet connection (broadband, 10 Mbps download or higher)
- Students can download most required software from the Rutgers University Software Portal. From the software portal, Mac and Windows users can download Microsoft Office, which includes PowerPoint. Class slides will be distributed in PowerPoint.
- Zoom. Zoom will be used if class must move online for some reason. It is also available to you if you are ill and cannot attend class in person. When using Zoom, use your RU Zoom account. Instructions for activating and signing in to a Rutgers Zoom Account are:
    - Instructions for Activating Your Rutgers Zoom Account
    - Instructions for Signing into Your Rutgers Zoom Account
- Technology Support
    - If you experience any technology issues, please contact Rutgers Business School Office of Technology and Instructional Services (OTIS), which offers extensive support coverage from 8:00 a.m. to 8:00 p.m., Monday through Friday
    - You can reach OTIS by emailing helpdesk@business.rutgers.edu

Financial Support Resources

- RBS Newark students in need of financial assistance may submit a request through the CARE Team form
- Students can also benefit from reviewing the Learning Remotely resource page

Disability Policy and Resources

If you need accommodation for a *disability*, obtain a Letter of Accommodation from the Office of Disability Services. The Office of Disability Services at Rutgers, The State University of New Jersey, provides student-centered and student-inclusive programming in compliance with the Americans with Disabilities Act of 1990, the Americans with Disabilities Act Amendments of 2008, Section 504 of the Rehabilitation Act of 1973, Section 508 of the Rehabilitation Act of 1998, and the New Jersey Law Against Discrimination. More information is available on the Rutgers Office of Disability Services Website.

- Rutgers University–Newark: Call (973) 353-5375 or email ods@newark.rutgers.edu

If you are experiencing a temporary condition or injury that is affecting your ability to fully participate in class, you should submit a request for support through the Rutgers Temporary Conditions Website.

Title IX Resources

If you are pregnant, the Office of Title IX and ADA Compliance is available to assist with any concerns or potential accommodations related to pregnancy.

- Rutgers University–Newark: Contact the Office of Title IX and ADA Compliance by phone at (973) 353-1906 or email TitleIX@newark.rutgers.edu

Religious Accommodations

If you seek religious accommodations, the Office of the Dean of Students is available to verify absences for religious observance, as needed.

- Rutgers University–Newark: Contact the Dean of Students Office at (973) 353-5063 or email deanofStudents@newark.rutgers.edu

VPVA and Harassment

If you have experienced any form of gender or sex-based discrimination or harassment, including sexual assault, sexual harassment, relationship violence, or stalking, the Office for Violence Prevention and Victim Assistance provides help and support. More information is available on the Rutgers Office for Violence Prevention and Victim Assistance Website (VPVA).

- Rutgers University–Newark: To report an incident, use the Rutgers Newark Incident Reporting Form. For support, you may contact the Office of Title IX and ADA Compliance at (973) 353-1906 or email TitleIX@newark.rutgers.edu. If you wish to speak with a confidential staff member who does not have a reporting responsibility, you may contact the Newark Office for Violence Prevention and Victim Assistance at (973) 353-1918 or email run.vpva@rutgers.edu

Bias Incidents

An act – either verbal, written, physical, or psychological that threatens or harms a person or group on the basis of actual or perceived race, religion, color, sex, age, sexual orientation, gender identity or expression, national origin, ancestry, disability, marital status, civil union status, domestic partnership status, atypical heredity or cellular blood trait, military service or veteran status.

- Report a Newark Bias Incident

Veteran and Military Services

If you are a military veteran or currently on active duty, you can obtain support through the Rutgers Office of Veteran and Military Programs and Services.

Mental and Physical Health Services

If you are in need of mental health services, please use our readily available services.

- Rutgers University–Newark Counseling Center

If you are in need of physical health services, please use our readily available services.

- Rutgers Health Services – Newark

Legal Support

If you are in need of legal assistance, please visit the Rutgers University Student Legal Services website to access support and resources.

Academic Support Services

Students experiencing difficulty in courses due to English as a second language (ESL) should contact the Program in American Language Studies for supports.

- Rutgers–Newark: PALS@newark.rutgers.edu

If you are in need of additional academic assistance, please use our readily available services.

- Rutgers University–Newark Learning Center
- Rutgers University–Newark Writing Center

Digital Accessibility Statement

Rutgers University is committed to ensuring that all digital course materials and technologies are accessible to every student. If you experience any difficulty accessing content used in this course, please contact me by email at *jo531@scarletmail.rutgers.edu* so that the necessary support can be provided.