



Big Data: Technologies and Applications

Prof. Il-Yeol Song, College of Computing & Informatics, Drexel University

Email: song@drexel.edu

Class: Morning session

SHORT COURSE DESCRIPTION

The era of big data has arrived. Big data brings us the data-driven paradigm and enlightens us to challenge different classes of problems we were not able to solve in the past. There are myriads of big data technologies available to us and much hype on how we can make the most of those technologies to extract values from big data projects. This course is a gentle introduction to big data for non-technical students with no programming knowledge required. The course will overview big data technologies and their applications to real-world projects. Topics to be covered include review of traditional relational database technologies with SQL, data-driven paradigm, big data warehousing and data lake, big data processing with Hadoop and Spark, cloud computing, NoSQL and NewSQL databases, visualization, artificial intelligence, machine learning and data mining algorithms, predictive analytics, internet of things (IOT), and blockchain. Examples of big data projects in smart healthcare areas and other innovative applications will be used to illustrate big data projects. Upon completing the course, students will be able to understand a broad spectrum of big data technologies and how the data-driven paradigm & big data technologies are used to create innovative applications and revolutionize our business environment as well as our society.

Requirements: This course will lightly cover several software such as SQL (for relational database queries), Tableau (for visualization), Watson Analytics (cloud-based automated data analytics), and MongoDB (NoSQL database). Students should be comfortable in learning and using them. It is expected that students bring their own laptop to class when they are covered.

INSTRUCTOR

Dr. Il-Yeol Song is professor in the College of Computing and Informatics of Drexel University, USA. His research interests include conceptual modeling, data warehousing, big data technologies & analytics, and smart aging. He published over 200+ papers in his research areas. Dr. Song served as Deputy Director of NSF-sponsored research center on Visual & Decision Informatics (CVDI) between 2012-2014. He was named an ACM Distinguished Scientist in 2013, and an ER Fellow in 2012. He is the recipient of 2015 Peter P. Chen Award in Conceptual Modeling. He won four teaching awards from Drexel – Exemplary Teaching Award (1992), Teaching Excellent Award (2000), Lindback Distinguished Teaching Award (2001), and the Most Outstanding Instructor Award (2011) from Drexel University. Dr. Song is a co-Editor-in-Chief of Journal of Computing Science and Engineering (JCSE) and a Consulting Editor for Data & Knowledge Engineering. He won the Best Paper Award in the IEEE CIBCB 2004. He delivered several keynote speeches on big data at the international conferences, including the First Asia-Pacific iSchool Conference in 2014, ACM SAC 2015 conference, ER2015 Conference, EDB 2016 Conference, A-LIEP 2016 Conference, and DEXA 2018 Conference.

READING MATERIALS

A syllabus with weekly study topics and reading materials is provided to the students. The lecture notes and reading materials will be available to download from the website.

COURSE REQUIREMENTS AND GRADING

Students are expected to attend all the classes. Grading components include mid-term exam, class attendance, and group project report or group project presentation. Due to a large number of students, each team can choose to either submit a written report in MS Word or make a presentation using a power point file in the last class. A written report could be around 10+ page long. A presentation could last about 15 minute with 20+ slides per team. Each student will join a team of four students to work on a collaborative project that studies a case study of an innovative Big Data application in any chosen domain area such as healthcare, business, environment, science, engineering, or social computing or learn a tool related to big data. All exams, reports, and presentations must use English. An attendance sheet will be available for students to sign up in each class. Exam questions will consist of true/false, multiple choices, and discussion questions. Grading will be computed as follows:

Exam :	40%
Attendance:	10%
Group project proposal	10%
Group project report or presentation:	40%

A grade of 60 or above is considered as a passing grade. Attendance in each class will be recorded.

- Pass (P): attendance of 80% or above (within 3 unexcused absence) and an average grade of 60 or higher.
- Fail (F): attendance below 80% (4 or more unexcused absence) or an average grade of lower than 60.

COURSE SCHEDULE

– WEEK I –

Thursday (27 June)

- Course overview
- Understanding Big Data (I)

Friday (28 June)

- Understanding Big Data (II)
- Evolution of Database Technologies
- Big Data and Data Science Career

– WEEK II –

Monday (1 July)

- Data-Driven Paradigm and Data Economy
- Big Data Use Cases
- **Recommended Project topics**

Tuesday (2 July)

- Relational databases
- SQL

Wednesday (3 July)

- Data Warehouses
- OLAP, OLTP, ETL
- Data Virtualization
- Data Lake

Thursday (4 July)

- Business Intelligence
- Data Science
- Data Analytics Lifecycle
- *Proposal Due*

– WEEK III –

Monday (8 July)

- Big Data Processing Architecture
- Hadoop and its Ecosystems
- Spark
- Cloud Computing

Tuesday (9 July)

- Visualization
- Tableau
- *Review for mid-term*

Wednesday (10 July)

- *Mid-term*
- Machine Learning and Data Mining (I)

Thursday (11 July)

- Machine Learning and Data Mining (II)
- Predictive Analytics

– WEEK IV –

Monday (15 July)

- Cloud-based automated Analytics
- Watson Analytics

Tuesday (16 July)

- NoSQL Databases
- NewSQL Databases
- MongoDB

Wednesday (17 July)

- Artificial Intelligence

- Internet of Things (IOT)
- Blockchain
- Review for Final Exam

Thursday (18 July)

- Big Data Trends, Opportunities, and Challenges
- Final Exam review

Friday (19 July)

- *Final exam*
- *Term Project Presentation*

IMPORTANT DATES (Subject to change)

- Mid-term Exam : Wednesday, July 10
 - Group Project topic discussion and teaming: Monday, July 1
 - Group Project proposal Due: Thursday, July 4
 - Group Project Report Due: Friday, July 19, 9:00PM
 - Group Project Presentation: Friday, July 19, 9:00AM
 - Final Exam: Friday, July 19, 11:00AM
- Email your report or PPT file to song@drexel.edu*