Big Data: Teachnologies 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 database technologies, data-driven paradigm, Hadoop, Spark, NoSQL and NewSQL databases, in-memory databases, cloud computing, big data warehousing, artificial intelligence, machine learning, cloud-based predictive analytics, internet of things (IOT), and blockchain. Examples of big data projects in smart city, smart healthcare areas, and other innovative applications will be used to illustrate big data projects. Students will be able to understand how the data-driven paradigm and big data technologies are used to create innovative projects and revolutionize our business environment as well as our society.

Requirements: This course will lightly cover SQL and Watson Analytics. 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. His research interests include conceptual modeling, data warehousing, big data management & analytics, and smart aging. Dr. Song served as Deputy Director of NSF-sponsored research center on Visual & Decision Informatics (CVDI) between 2012-2014. He is an ACM Distinguished Scientist and an ER Fellow. 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). 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, and A-LIEP 2016 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 ISS course website (http://summer.skku.edu).

COURSE REQUIREMENTS AND GRADING

Students are expected to attend all the classes. Grading components include mid-term exam, final exam, class attendance, and group project report or group project presentation. All exams, reports, and presentations must use English. Due to a large number of students, each team can choose to either submit a written report in 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 minutes per team. An attendance sheet will be circulated for students to sign up in each class. Each student will join a team 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. Grading will be computed as follows:

Exam 1:	25%
Exam 2:	25%
Attendance:	10%
Group project proposal	10%
Group project report or presentation:	30%

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: Introduction to Big Data –

Tuesday (26 June)

- Course overview
- Understanding Big Data (I)

Wednesday (27 June)

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

Thursday (28 June)

- Data-Driven Paradigm and Data Economy
- Big Data Use Cases

Friday (29 June)

- Relational databases
- SQL (I)

– WEEK II: Big Data Technologies and Applications (I) –

Monday (2 July)

- SQL (II)

Tuesday (3 July)

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

Wednesday (4 July)

- Business Intelligence
- Data Science
- Data Analytics Llfecycle
- Review for Mid-Term

Thursday (5 July)

- Mid-Term
- Big Data Processing Architecture
- Hadoop and its Ecosystems
- Spark

– WEEK III : Big Data Technologies and Applications (II) –

Monday (9 July)

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

Tuesday (10 July)

- Machine Learning and Data Mining (II)

Wednesday (11 July)

- Cloud-based Analytics
- Watson Analytics

Thursday (12 July)

- NoSQL Databases
- NewSQL Databases

– WEEK IV: Wrap-up–

Monday (16 July)

- In-Memory Databases
- Cloud Computing
- Artificial Intelligence
- Internet of Things (IOT)

Tuesday (17 July)

- Smart health
- Smart city
- Blockchain
- Big Data Trends, Opportunities, and Challenges

- Final Exam and Review

<u>Wednesday (18 July)</u>

- Final Exam
- Term Project Presentation