Dongseo University

Division of Computer Engineering

Big Data Processing

Instructor(s): Hyontai Sug, Ph.D.   **Classroom: UIT308**

Office: UIT309   **Class Time: MWF 10:00AM-10:50AM**

Phone: 051-320-1733   **Office Hours: Tue~Th, 1:00PM-3:00PM**

Email: sht@gdsu.dongseo.ac.kr

Course Description:

This course introduces principles of big data processing technology including related algorithms, systems, and advanced topics related to large-scale data management and knowledge extraction techniques. The course will cover big data models, databases and query languages, large scale distributed database systems, data mining algorithms, and big data applications. Implementation of a distributed database on a standalone machine will be covered and students will learn how to build their own database for big data. The course will also cover critical topics in mining and knowledge discovery of big data with various applications.

Course Goals & Objectives:

At the conclusion of this course, the successful students will be able to

- Describe big data concepts
- Create big data systems using popular database management system tools
- Solve big data problems by constructing related systems
- Develop insights into data mining and knowledge discovery of big data
- Evaluate the effectiveness of big data systems in computer information systems

Course Outline:

- **Week 1 : introduction to big data technology**

- **Week 2 : relational data models and semi-structured data**

- **Week 3 : graph theory and graph data models**

- **Week 4 : knowledge discovery process**

- **Week 5 : relational databases**

- **Week 6 : beyond relational databases**

- **Week 7 : extracting information from big data 1**

- **Week 8 : Mid-Term Week**

- **Week 9 : extracting information from big data 2**

- **Week 10 : scalable big data processing 1**

- **Week 11 : scalable big data processing 2**

- **Week 12 : data to knowledge 1**

- **Week 13 : data to knowledge 2**

- **Week 14 : term project discussion**

- **Week 15 : Final Week**

Textbook(s)

- Required: Seven Databases in Seven Weeks, E. Redmond, J.R. Wilson, Pragmatic Bookshelf, 2012.
- Recommended: Data Science for Business, F. Provost, T. Fawcett, O'Reilly Media, 2013.

Class Website: e-Class

Course Assignments & Grading:

- *Exams: there will be two exams; mid-term and final exam.*

- *Quizzes*: quizzes may be given if necessary.

- *Course Projects*: there will be a term project that will cover big data processing

- *Assignments*: there will be one or two assignments related to survey and proposal of big data system development.

- *Bonus Credit*: students who are actively participate the class may get extra points

- *Grading*: mid-term exam, final exam, term project 25% each, attitude 5%, homework and presence 10%

Grading Policies:

• *Late Work:* All assignments must be submitted on the due date. Late assignments will not be accepted.

• *Make-Ups:* There will be no make-ups given for any of the assignments in this course.

• *Contesting:* Grades can be contested during a two-week period from the time that they are announced. After such period has elapsed, grades may not be contested.

Course Policies:

- Attendance: class attendance will be counted and will be 10% of your grade
- Academic Misconduct Policy: We really do not expect it, so please do not disappoint us! However, any form of cheating will be penalized and may result in failing the course or expulsion from the university