



School of Information Sciences

College of Communication & Information

**IS 592 Big Data Analytics
Spring 2015**

Instructor: Peiling Wang, Ph.D.
Class held: Thursday 11:10 am - 1:55 pm in Blackboard Collaborate
Office: 443 Communications Building; voice mail: (865) 974-3700
Hours: Wednesday 1:30 - 5 p.m.; also available by appointment
Contact: peilingw@utk.edu

Catalog Description

Introduces the concepts big data and data analytics as an emerging field. To address the opportunities and challenges of big data in academics, businesses, sciences, the Web, etc. To understand the nature of big data analytics and their various contexts. To master basic concepts and process of data analytics. To design analytics initiatives/proposals. To practice data mining techniques and skills. To explore data modeling and visualizing.

Pre-requisite: Database Management Systems (completion of IS584 or equivalent)

Goals/Objectives

- To survey the needs and importance of data analytics in various contexts
- To understand the challenges of managing big data
- To practice data extraction, transformation and load techniques (ETL)
- To develop algorithms to analyze and model data
- To design effective ways for communicating results to special users

Methods of Teaching/Learning

This course is built on knowledge and skills of database management systems, research, and statistics. Big data and its analysis are about real world applications. The focus will be on issues challenging organizational decision-making, real world data needs that call for methods of data management, analytics, and modeling to derive new knowledge for better decision making. Students are expected to read broadly and to work on real data sets from various contexts. This course is managed using Blackboard Learn courseware, which is accessible using your UT NetID and Password at <https://bblearn.utk.edu/>. The Blackboard Collaborate, a tool hosted in Blackboard, will be used for **synchronously** for virtual class sessions; you may attend classes from anywhere in the world. The course materials, tasks, and grades are accessible in Blackboard Learn.

To complete hands on tasks, you must have access to a DBMS system such as ACCESS, SQL server, Oracle, MySQL, etc. If you have programming background, you may alternatively extract data by programming. To perform statistical analysis, you should use SPSS, SAS or similar tools. The alternative will be Excel that support basic statistical functions and good graphics.

Course Materials

Readings are loaded in Blackboard Learn.

Tasks and Evaluation Criteria

- **Attendance & Participation (10%)**

Prepared attendance and participation in course activities are important to success in this course. If you have to miss a class for whatever reasons, you are still responsible for the material covered. If you miss a class, you may replay the recording. Blackboard Collaborate keeps track of attendance and replay.

Class activities include presentations and discussion.

- **ePortfolio or Journal (10%)**

Be a reflective learner! Throughout the semester, you should maintain a learning journal or ePortfolio. Write journal entries to reflect your thoughts, analyze critical incidents, and check milestones.

If you have taken the ePortfolio course, you should continue building your ePortfolio in this course by writing Posts to reflect on your learning and achievements. At the end of the semester, you will write a reflective summary for the course as a Page in your ePortfolio.

If you have not taken the ePortfolio course, you may keep a structured journal with dated entries and write a final reflection piece. You submit the reflection along with selected journal entries in any format accessible to the instructor.

Make your learning and achievements visible through the development of a course ePortfolio. Journal entries or ePortfolio Posts document your learning and professional growth with evidence and through reflection on learning experiences. Both collecting artifacts and reflecting in journal entries are private actions but presenting outcomes and sharing reflective summary are oriented toward a product for public (or your evaluators).

What to write in journal entries (ePortfolio posts)? You do not need to report or log what you have done during the course. You need to focus on significant learning incidents, aha moments, relevant thoughts, analysis and synthesis of important concepts, and milestone checking. Reflection is a higher level of cognitive activity in which you makes sense of what and how you learned. For example, when you encountered a challenging problem, you should reflect on the strategies and the process through which you were, or were not, able to solve the problem. For ePortfolio students, you should classify your journal entries so that they can be easily accessed to facilitate a higher level of synthesis later in producing your final ePortfolio. For non ePortfolio students, you should structure your journal with meaningful headings, which will help you to develop a summary reflection of the semester as your last journal entry.

- **Tasks (Due dates are posted):**

1. Self-Efficacy of Course Preparedness

This survey is a part of course participation

2. *Data Science and Data Scientists (15%)*

Understand the nature of data analytics in context. Understand the skill set of data scientists.

3. *Data Preparation: Extract, Transform and Load (ETL) (25%)*

Extract the relevant data from original sources (the raw data); transform raw data to appropriate

format; load the transformed data to a database.

4. *Data Analysis* (25%)

Explore the transformed data to derive meaningful results (statistical analysis, pattern recognition, trend visualization)

5. *Reporting* (15%)

Write one report for your end users of your results such as CEO, president, Deans, Directors, ...

Write a document for your peer professionals who might join you in the project or continue the work you have completed

On Grading:

The University of Tennessee grading system for graduate level courses are as follows. However, grades, as a quantitative assessment of course performance, do not always reflect a student's real competences.

A: superior performance

B+: better than satisfactory performance; **B:** satisfactory performance

C+: less than satisfactory performance; **C:** performance well below the standard expected of graduate students

D: clearly unsatisfactory performance and cannot be used to satisfy degree requirements

F: extremely unsatisfactory performance and cannot be used to satisfy degree requirements

Academic Integrity

"The responsibility for learning is an individual matter. Study, preparation and presentation should involve at all times the student's own work, unless it has been clearly specified that work is to be a team effort. Academic honesty requires that all work presented be the student's own work, not only on tests, but in themes, papers, homework, and class presentation. ..." (*Hilltopics Student Handbook*, The University of Tennessee, Knoxville; <http://dos.utk.edu/publications/hilltopics/index.shtml>).

Cheating, plagiarism, providing unauthorized help and other acts of dishonesty violate the rule of academic honesty; the offender will be subject to penalties as set forth in *Hilltopics*

Special Needs

If you need course adaptations or accommodations because of a documented disability or if you have emergency information to share, please contact the Office of Disability Service at 191 Hoskins Library or at (865) 974-6087. This will ensure that you are properly registered for services.

Diversity Statement

CCI recognizes and values diversity. Exposing students to diverse people, ideas and cultures increases opportunities for intellectual inquiry, encourages critical thinking, and enhances communication and information competence. When all viewpoints are heard, thoughtfully considered, and respectfully responded to, everyone benefits. Diversity and fairness unite us with the wider professional and global community.

Policy on Inclement Weather

If the University Knoxville campus is officially closed, classes will be canceled.

IT Problems

In case of Internet or Blackboard issues, call OIT at (865) 974-3117 or SIS DE support at 888 378 9338 or 865 974 7913.

Notes: In most cases, Blackboard Learn will have announcements regarding special arrangements. However, in case of IT or weather issues that may prevent effective delivery of announcements, I can send cell phone text messages. If you wish to receive these messages, please text me with your name (I will provide my phone number in class).

Schedule: (Subject to revision due to unforeseen circumstance)

<i>January 8</i>	Welcome & Introduction
<i>January 15</i>	What is data science? What skills are required for data scientists?
<i>January 22</i>	What is big data? What is big data analytics? Basics for Data Analytics
<i>January 29</i>	Data and Data Model (Review of ERD)
<i>February 5</i>	<i>Student presentation on selected topics</i>
<i>February 12</i>	Data Preparation (ETL)
<i>February 19</i>	Research: from Questions to Answers; Collecting Facts and Measuring Concepts
<i>February 26</i>	Statistical Tools for Data Analytics
<i>March 5</i>	Combine Knowledge and Skills from Database and Statistics
<i>March 12</i>	Big Data (From Database to Data Warehouse)
<i>March 26</i>	Guest speaker
<i>April 2</i>	<i>Methods and Tools for Structured Data</i>
<i>April 9</i>	<i>Methods and Tools for Unstructured Data</i>
<i>April 16</i>	Visualization of Analytical Results
<i>April 23</i>	<i>Student Presentation</i>