# Fall 2013 Course (Undergrad/Grad)
# CSE 489-02 / CSE 589-03 – Advances in Big Data Management

**Instructor: A. Rezgui Email: rezgui@cs.nmt.edu Office: Cramer 224**

**Class Meeting Time:** Monday/Wednesday, 2:00pm-3:15pm (Cramer 227)
**Objective:** The objective of this course is to explore recent advances in big data management through an in-depth study of authoritative references and recent literature. The course is offered for undergrad and grad students.
**Assignments:** Students will have to spend a significant amount of time reading recent papers published in high impact journals or presented at major Big Data conferences, e.g., the IEEE Conference on Big Data (BigData), the IEEE International Conference in Big Data Science & Engineering (BSDE), the ASE/IEEE International Conference on Big Data, and the International Conference on Big Data Analytics (BDA).

While **undergrad students will not be subject** to the same grading scheme as grad students, the assignments and requirements are the same. They include:

1. **Presentations:** Each student will have to make three presentations, each covering a topic in Big Data management. The presentations could be of practical nature (e.g., an introduction to the Hadoop distributed file system (HDFS)) or of higher level and based on one or more research papers. Students may select the topics of these presentations or I may assign a paper/topic.
2. **Project:** Each student will work on a semester-long project related to Big Data. Students may select the topic of their projects or I may assign a topic. A project could be, for example:
   - the evaluation/comparison of several approaches published in the literature,
   - identifying a problem in Big Data management and designing/implementing/evaluating a solution to that problem,
   - exploring a new idea in Big Data management
   - improving an existing algorithm/technique and comparing the improvement with the original algorithm/technique.

Each student will have to meet with the instructor every two weeks to report on his/her progress. Students will have to submit a report for their project by the end of the semester. They also will have to submit a first draft of that report (with about 50% of the work done) in the middle of the semester.

**Organization & Content:** Typically, lectures will be tutorial in nature, i.e., introducing topics that are then explored more in-depth by studying content from other sources. In addition to the lectures, some classes will be for presentations by students.

**Grades:** Grades will be based on the following distribution:

| | |
|---|---|
| Participation: | 20% |
| Presentations: | 30% |
| Project (Report/Paper and Presentation): | 50% |

Topics covered will be selected from a spectrum of topics that includes (but is not limited to):

# Fall 2013 Course (Undergrad/Grad)
# CSE 489-02 / CSE 589-03 – Advances in Big Data Management

**Instructor: A. Rezgui  Email: rezgui@cs.nmt.edu  Office: Cramer 224**

**Introduction to Big Data**
- What is Big Data?
- The four dimensions of Big Data: volume, velocity, variety, veracity
- Why is Big Data important?

**Hadoop**
- History
- Components of Hadoop
  - The Hadoop Distributed File System (HDFS)
  - An introduction to MapReduce
  - Hadoop Common Components
- Application Development in Hadoop
  - Pig and Pig Latin
  - Hive
  - Jaql
- Getting data into Hadoop
  - Basic Copy Data
  - Flume
- Other Hadoop Components
  - Oozie

**Other Big Data stores**
- An introduction to NoSQL Databases
  - Key value / tuple Stores
    - DynamoDB
    - Riak
    - Redis
  - Column Stores
    - Hadoop/HBase
    - Cassandra
    - Hypertable
  - Document stores
    - MongoDB
    - CouchDB
  - Graph Stores
    - Neo4J
    - InfiniteGraph

**Processing Big Data**

**Analyzing Big Data**

**Other Topics From Conference/Journal Papers**