

CS 6500 : BIG DATA ANALYTICS

<i>Semester Hours:</i>	3.0	<i>Contact Hours:</i> 3
<i>Coordinator</i>	Robert Green	
<i>Text</i>	Selected research publications	
<i>Authors:</i>	VARIOUS	
<i>Year</i>	Various	

SPECIFIC COURSE INFORMATION

Catalog Description:

This course focuses on analytic tools for unstructured data. The students will learn basic and advanced analytical methods. Big data analytics technology and tools, including Hadoop and MapReduce, will be introduced. It also discusses SQL extensions and other advanced SQL techniques for in-database analytics. Prerequisite: CS 5620 or equivalent. Approved for distance learning education. Credit cannot be earned for both CS 6500 and MSA 6500.

Course type: **ELECTIVE**

SPECIFIC COURSE GOALS

- I can explain Big Data concerns in the field.
- I can define the role of a data scientist.
- I can describe the life cycle of data analytics.
- I can save and retrieve ultra large scale data.
- I can use parallel processing to scale-up applications and speed up their execution.
- I can understand Hadoop-like distributed computing frameworks and the tools available.
- I can use Hadoop and/or other distributed file systems to store vast quantity of data.
- I can write and execute MapReduce programs to analyze unstructured data.
- I can utilize tools to monitor the health and performance of Hadoop-like clusters.
- I can describe the advantages and disadvantages of using NoSQL databases for Big Data.

LIST OF TOPICS COVERED

1. Overview (~10%)
2. Distributed Filesystems (~10%)
3. Map-reduce Programs / Hadoop (~10%)
4. Dataflow Languages / Spark (~10%)
5. Column-oriented Storage Techniques (~10%)
6. Interactive Analysis (~10%)
7. NoSQL (~10%)
8. Amazon Web Services (~10%)
9. Wide Column Stores / HBase (~10%)