1. **Course number and name**
   CptS 415: Big Data

2. **Credits and contact hours**
   3 credits, 3 lecture hours

3. **Instructor’s or course coordinator’s name**
   Yinghui Wu

4. **Textbook, title, author, and year**

   **Other supplemental materials**
   - *Big Data: A Revolution That Will Transform How We Live, Work, and Think*  
     <http://www.amazon.com/BigData-Revolution-Transform-Think/dp/0544227751>

5. **Specific course information**
   a. **Catalog description:** Big data models, databases and query languages, modern distributed database systems and algorithms. (Crosslisted course offered as CPT S 415, CS 415)
   b. **Prerequisites or corequisites:** CPT-S 215, 223, or 233, with a C or better.

6. **Specific goals for the course**
   By the end of the course, students will be able to
   - Display comprehensive understanding of basic and core concepts in Big data models (the Big V’s, Relational, NoSQL, semi-structured, networks) and systems (2a-2d)
   - Understand principles of system and algorithm design to cope with Big Data challenge (Index, Sampling, Data Compression, Views, Approximate Query) (1a, 1d, 1e, 2b-2e)
   - Understand principles of cost analysis and complexity analysis for Big Data applications (6a, 6b, 6c, 6d)
   - Develop algorithms and prototype systems using (open-source) Big data platforms (NoSQL, MapReduce/Hadoop, Spark) and software packages with emphasize on real-world social impact (2b, 2c, 2d, 2f, 2g, 4e, 4f)
   - Experience effective scientific project presentation, writing and experimental study (3a, 3b, 3c, 3d, 3e, 3f).
• Get familiar with research frontier (Data quality, privacy, security, ethics) in academic and industrial community related to core Big Data management techniques (4a, 4c)
• Work effectively in a Big Data project team (5b, 5c, 5d, 5e, 5g, 5f)
• Effectively manage team projects (5a, 5c, 5d).
• Effectively identify and exploit resources and literatures of Big Data for course project (7a-7d)

7. **Brief list of topics to be covered**
   • Big Data models, relational DBMS
   • Beyond Relational data: XML and RDF
   • NoSQL and NewSQL
   • Approximate Query Processing
   • Distributed and parallel models
   • Scalable search: MapReduce and Hadoop
   • Cloud computing
   • Data mining
   • Data quality
   • Data Privacy and Ethics