CS 6307

Big Data Management & Analytics for non-CS Majors

Anurag Nagar, Ph.D.
Computer Science
University of Texas at Dallas

Topics

- Focus on databases - traditional and NoSQL

- Big Data basics

- Big Data storage

- Analytics on large datasets

- Data Mining of large datasets

- Graph Analytics

Topics

- Theory followed by lab sessions.

- Various computational tools will be used. You will need to install these on your local computers.

- Need to be familiar with your own computer - installation, setting path, running Linux commands.

- Need to know at least one modern programming language e.g. Python, Java, .NET, etc

Software

- Database part:
  Install Anaconda with ipython add-in
  https://anaconda.org/conda-forge/ipython-sql

- NoSQL part:
  MongoDB cloud, Amazon cloud, Google cloud

- Analytics part:
  Apache Spark

If you have trouble installing software, the TA can help you.

Please apply for free credit on AWS Educate:
https://aws.amazon.com/education/awseducate/apply/

Administrative Issues

- Read the syllabus very carefully.

- Review the grading details.

- Take the course policy quiz. Make sure you understand each point.

- Keep up with the material.

- If at any point, you feel you are lagging behind please see instructor.