CMP4011
Big Data and Cloud Computing

Course Information

Dr. Lydia Wahid

Instructor Info

- *Course Instructor:* Dr. Lydia Wahid
- *Email:* LydiaWahid@Outlook.com
- *Office hours:* Tuesday 2-4
  Wednesday 1-2

Objectives of the Course

- Learn the concepts of Data Mining, Big Data, and Data Analytics.
- Learn Big Data and Big Data Analytics techniques and how to solve problems whose solutions are enabled by technology that can support the analysis of Big Data datasets.
- Learn the basics of cloud computing and use cloud computing providers to develop big data applications.
- Acquire the necessary knowledge to develop working code for real-world Big Data applications.

Learning Outcomes (LOs)

1. Apply MapReduce processing technique and Design its different components to suit the Big Data problem at hand. (NARS mapping: C.1)
2. Use the different Big Data processing frameworks to build a Big Data application. (NARS mapping: B.3)
3. Analyze and Apply Big Data Predictive and Descriptive Analytics. (NARS mapping: C.2)
4. Use cloud computing providers such as Microsoft Azure and Amazon Web services (AWS) (NARS mapping: C.2)
5. Describe data warehousing and the different storage technologies. (NARS mapping: A.3)
6. Develop a hands-on experience by developing a real-world Big Data application through the term project. (NARS mapping: C.1)

Grades Distribution

- Two-Semester System:
  - Final Exam → 60 marks
  - Midterm Exam → 10 marks
  - Labs → 10 marks
  - Project → 15 marks
  - Class Assignment → 5 marks

  - Total → 100 marks

Attendance

- The attendance will be recorded each week normally by calling students names during Lecture time and Section time.

Exams Policy

- Midterm and Final exams will be **Open-Book**:
  - You can only get hardcopy material.
  - **No** electronic devices are allowed.
  - You can **only** get the following material: Lecture slides, sheets, and your own written notes (not photocopied nor printed on PC)
  - **No** borrowing of any material from your colleagues during exam time is allowed.

References – Big Data Analytics

- EMC Education Services. *Data science and big data analytics: discovering, analyzing, visualizing and presenting data*. Wiley, 2015.
- C.S.R., P., et al. *Big Data Analytics: Systems, Algorithms, Applications*. Springer Nature, 2019.
- Schintler, L. and McNeely, C. *Encyclopedia of Big Data*. Springer International Publishing, 2020.

References – Cloud Computing

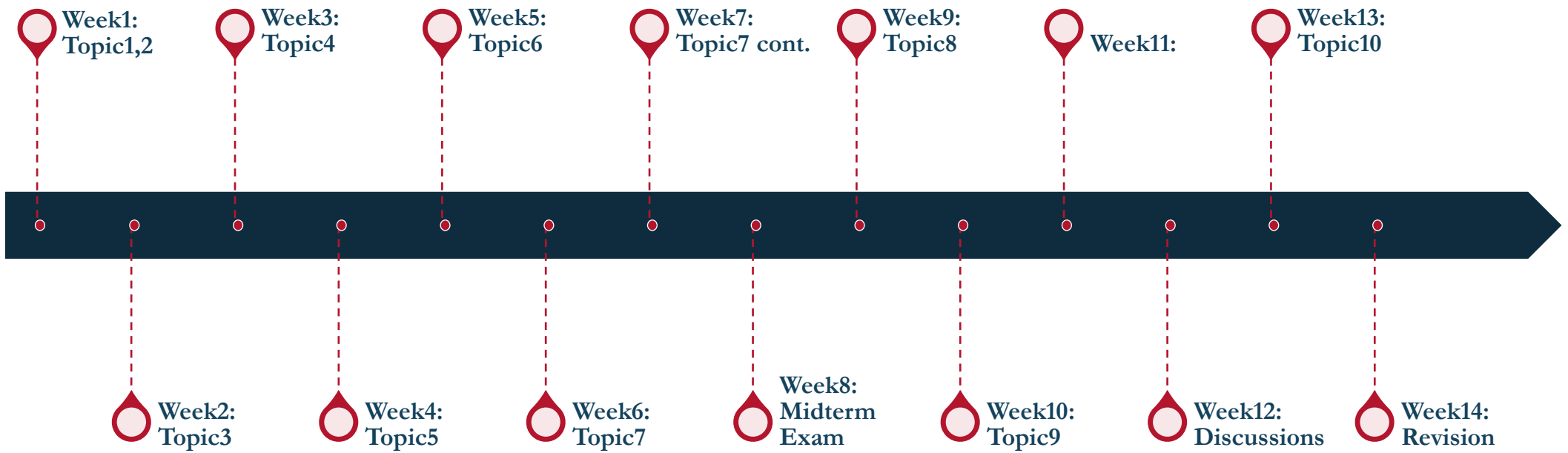
- Cloud Computing: Concepts, Technology & Architecture by Thomas Erl et al. Prentice Hall, 2013.

- Introduction to Cloud Computing by Eric Frick 2021.

Topics

1. Fundamentals of Big Data
2. Fundamentals of Data Analytics
3. Big Data Processing Techniques
4. Big Data Processing Frameworks
5. Big Data Analytics – Predictive Analytics
6. Big Data Analytics – Descriptive Analytics
7. Cloud Computing
8. Data Warehouse
9. Big Data storage concepts and strategies
10. Big Data storage technologies

Lectures Schedule



Lectures Schedule

Week1:	Fundamentals of Big Data and Data Analytics
Week2:	Big Data Processing Techniques
Week3:	Big Data Processing Frameworks
Week4:	Big Data Analytics – Predictive Analytics
Week5:	Big Data Analytics – Descriptive Analytics
Week6:	Cloud Computing
Week7:	Cloud Computing
Week8:	Midterm Exam (to be confirmed later)
Week9:	Data Warehouse
Week10:	Big Data storage concepts and strategies
Week11:	
Week12:	Class Assignment Discussion
Week13:	Big Data storage technologies

Sections Schedule

Week1:	Lab1 – Introduction to R
Week2:	Lab2 – Hadoop Framework
Week3:	Lab3 – Integrating Hadoop and R
Week4:	Lab4 – Apache Spark Framework
Week5:	Lab5 – Predictive Analytics
Week6:	Lab6 – Descriptive Analytics
Week7:	Lab7 – Cloud Computing – Microsoft Azure
Week8:	Midterm Exam (to be confirmed later)
Week9:	Lab8 – Cloud Computing – Amazon Web Services (AWS)
Week10:	To be decided
Week11:	
Week12:	Sheet
Week13:	Final Project Delivery

Project

Week6:	**Project Proposal Delivery**
Week7:	**Project Proposal Feedback from the TA**
Week13:	**Final Project Delivery**

Note:

- Project Proposal :
  - All teams will receive feedback from the TA within 1 week from the delivery date.

Thank You