Fundamental of big Data Analytics

Instructor Name: Isbah Khalid   **TA Name (if any):**
Email address: asbah.khalid@nu.edu.pk **Office Location/Number:** New building room 19
Office Hours:

Course Information Program: BSDS
Credit Hours: 3
Core Prerequisites (if any): Programming competence, Intro to Data Science
Course Website (if any) : Google classroom
Class Meeting Time:

Textbook(s) /Supplementary Readings:
There is no standard one "textbook" for this course. The following book will be used as a primary text to guide some of the discussions, but it will be heavily supplemented with lecture notes and reading assignments from other sources.
- Mining of Massive Datasets by Jure Leskovec Stanford Univ. Anand Rajaraman Milliway Labs , Jeffrey D. Ullman Stanford Univ.

Additional references and books related to the course:
- White, Tom. "Hadoop: The definitive guide." O'Reilly Media, Inc., 2012. 2.
  Karau, Holden, Andy Konwinski, Patrick Wendell, and Matei Zaharia. "Learning spark: lightning-fast big data analysis." O'Reilly Media, Inc., 2015. 3.
- Miner, Donald, and Adam Shook. "MapReduce design patterns: building effective algorithms and analytics for Hadoop and other systems." O'Reilly Media, Inc., 2012.

Tentative Course Outline

Week	lecture	Topics
1	1	- What is Big Data and Why we need Big data - Sources of Data - Intro to big Data tools, - Data Engineering - Types of Data
	2	- Hadoop as a Platform - Hadoop Distributed File Systems (HDFS)
2		MapReduce Framework

3-4		Sql queries ● Simple ● Aggregated function and grouping ● Join ● Nested queries
5	Mid 1	
6		Hive, pig
7-8	1	Resilient Distributed Datasets (RDD)
	2	Apache Spark, Apache Spark SQL
9		Nosql Hbase , cassandra kafka
10	Mid 2	
11-12		Resource Management in the cluster (YARN)
13-14		Machine learning on Hadoop / Spark
15		Spark Streaming, Other Components of Hadoop Ecosystem
Final		

(Tentative) Grading Criteria
Quizzes 10%
 Assignments/Homeworks/Project 15 - 25%
Midterms 25%
Final Exam 45%
Total: 100