CSC 3303 BIG DATA ANALYTICS

DR. RAINI HASSAN

Department of Computer Science, KICT, IIUM

Office Room: C2-14 (Block C, Level 2)

Email: hrai@iium.edu.my

Office Telephone No: +603 6196 5655

Dua Before Studying

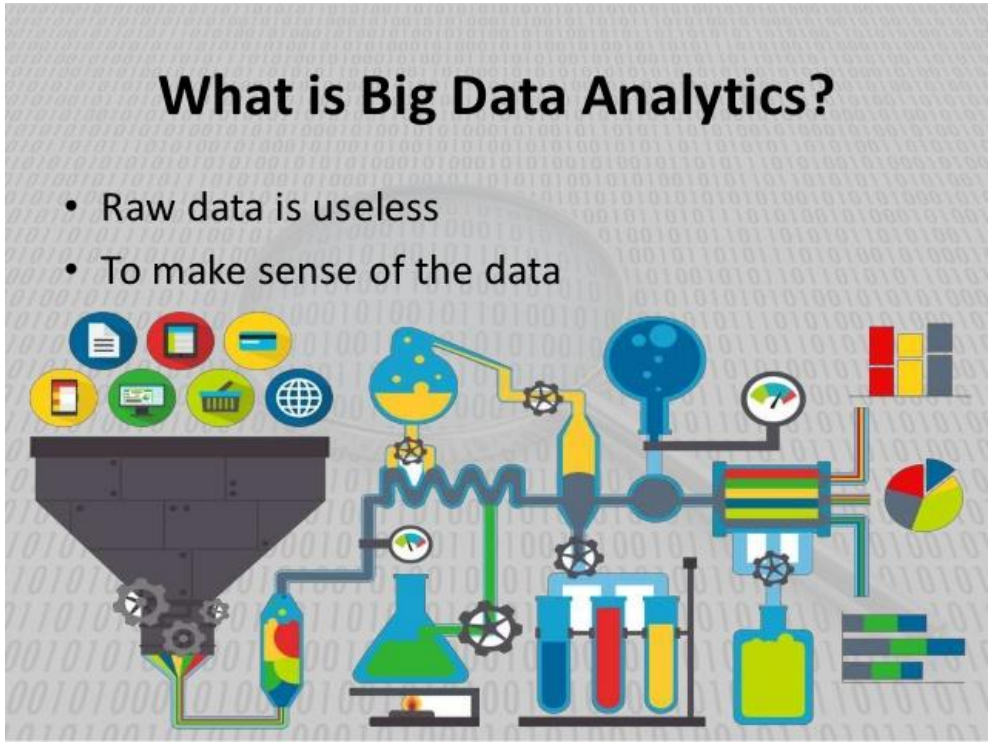
Dua Before Studying

Oh Allah! Make useful for me what you have taught me and teach me knowledge that will be useful to me. Oh Allah! I ask you for the understanding of the prophets and the memory of the messengers, and those nearest to you.Oh Allah! Make my tongue full of your remembrance and my heart with awe of you. Oh Allah! You do whatever you wish, and you are my availer and protector and best of aid.

اللَّهُمَّ انْفَعْنِي بِمَا عَلَّمْتَنِي وَ عَلِّمْنِي مَا يَنْفَعُنِي
اللَّهُمَّ إِنِّي أَسْأَلُكَ فَهْمَ النّ بِيِّينَ وَ حِفْظَ الْمُرْسَلِينَ الْمُقَرَّبِينَ.
اللَّهُمَّ اجْعَلْ لِسَانِي عَامِرًا بِذِكْرِكَ وَ قَلْبِي بِخَشْيَتِك. .
إِنَّكَ عَلَى مَا تَشَاءُ قَدِيرٌ وَ أَنْتَ حَسْبُنَا اللَّهُ وَ نِعْمَ الْوَكِيلُ.

Allahumma infa'nii bimaa 'allamtanii wa'allimnii maa yanfa'uunii. Allahumma inii as'aluka fahmal-nabiyyen wa hifzal mursaleen al-muqarrabeen. Allahumma ijal leesanee 'aiman bi dhikrika wa qalbi bi khashyatika. Innaka 'ala ma-tasha'u qadeer wa anta hasbun-allahu wa na'mal wakeel.

Welcome to the Big Data Analytics Course!



Outlines

1. Course Synopsis
2. Learning Outcomes
3. Assessment
4. Reference (Required Text)
5. Reference (Recommended Text)
6. Course Contents
7. My Timetable
8. Course Timetable
9. Additional Notes

1. Course Synopsis [1]

- Big data consist of mountains of mostly unstructured data in the form of:
  - web logs,
  - videos,
  - speech recordings,
  - photographs,
  - e-mails, and
  - tweets.

1. Course Synopsis [2]

- Big Data Analytics (BDA) involves the **process** of:
  - collecting,
  - storing,
  - organizing, and
  - analyzing large sets of data to discover patterns…
- that enables organizations to **better understand** a mix of:
  - structured,
  - semi-structured and
  - unstructured data…
- in search of valuable information and insights.

1. Course Synopsis [3]

- This course provides a detailed overview of big data and Data Analytics Lifecycle (DAL) to address application issues in both workplaces and research environments that leverage on big data.
- Students will be exposed to grounding basic analytic methods as well as standard big data analytics technology, method, tools and framework such as R, MapReduce and Hadoop.
- This course includes a capstone project involving the design of highly scalable big data systems that applies the concepts in the Data Analytics Lifecycle (DAL).

2. Learning Outcomes

1. To identify and explain the big data concepts and principles.
2. To apply Data Analytics Lifecycle (DAL) in problem solving and implement the solutions through big data tools.
3. To analyze, develop and evaluate a capstone project of highly scalable big data systems, that also applies the concepts of Data Analytics Lifecycle (DAL), effectively and professionally in a group work.

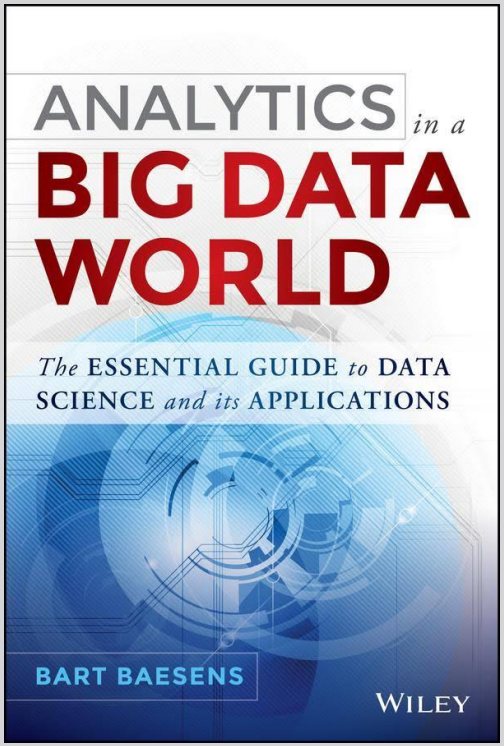
3. Assessment

Methods	Percentage
Quizzes	2 x 10% = **20%**
Assignments	2 x 10% = **20%**
Group Project	Proposal 5% Report 10% Presentation 5% = **20%**
Tests	2 x 20% = **40%**
TOTAL	**100%**

Project (20%) and Test 2 (20%) are the final assessment (total = 40%). The marks will be hidden from you.

4. Reference (Required Text)

- Baesens, B. (2014). *Analytics in a big data world: the essential guide to data science and its applications.* New Jersey, USA: John Wiley & Sons



5. Reference (Recommended Text)

- Camm, J. D., Cochran, J. J., Fry, M. J., Ohlmann, J. W., Anderson, D. R., Sweeney, D. J. & Williams, T. A. (2014). *Essentials of business analytics*. Connecticut, USA: Cengage Learning.
- EMC Education Services (2015). *Data science and big data analytics: discovering, analyzing, visualizing and presenting data*. Indiana, USA: John Wiley & Sons.
- Minelli, M., Chambers, M. & Dhiraj, A. (2013). *Big data, big analytics: emerging business intelligence and analytic trends for today's businesses*. New Jersey, USA: John Wiley & Sons.

6. Course Contents (1)

- All course materials such as .pdf lecture slides, assignments, quizzes, etc. can be accessed at Google Classroom.

https://classroom.google.com

- To join my class, use this class code:

84j1cy

6. Course Contents [2]

Lecture 1: Introduction to Big Data Analytics

- Big Data Overview
- State of Practice in Analytics
- Key Roles for the New Big Data Ecosystem
- Examples of Big Data Analytics

Main references:

1. Baesens, B. (ed) (2012) Data collection, sampling, and preprocessing, in analytics in a big data world: The essential guide to data science and its applications. John Wiley & Sons.
2. EMC Education Services (2015). Data science and big data analytics: Discovering, analyzing, visualizing and presenting data. John Wiley & Sons.

6. Course Contents [3]

Lecture 2: Collection, Sampling and Pre-Processing

- Introduction
- Types of Data Sources
- Sampling
- Types of Data Elements
- Visual Data Exploration
- Missing Values
- Outlier Detection and Treatment
- Variable Selection
- Standardizing Data
- Categorization
- Weights of Evidence Coding
- Segmentation

Main reference:

Baesens, B. (ed) (2012) Data collection, sampling, and preprocessing, in analytics in a big data world: The essential guide to data science and its applications. John Wiley & Sons.

6. Course Contents [4]

Lecture 3: Data Analytics Lifecycle (DAL)

- Phase 1: Discovery
- Phase 2: Data Preparation
- Phase 3: Model Planning
- Phase 4: Model Building
- Phase 5: Communicate Results
- Phase 6: Operationalize

Main reference:

EMC Education Services (2015). Data science and big data analytics: Discovering, analyzing, visualizing and presenting data. John Wiley & Sons.

6. Course Contents [5]

Lecture 4: Introduction to Parallel Processing/Programming

- Basic Concepts
- Big Data and Parallel Processing Architecture
- Machine Learning Meets Massively Parallel Processing
- Hadoop and Parallel Processing

References:
From multiple youtube videos.

6. Course Contents [6]

Lecture 5: Hadoop, HDFS & MapReduce

- A Brief History of Hadoop
- What is Hadoop?
- Installation
- Learn HDFS
- MapReduce
- Hadoop Example
- Counters & Joins in MapReduce

Main reference:
https://www.guru99.com/what-is-big-data.html

6. Course Contents [7]

Lecture 6: Predictive Analytics

- Target Definition
- Linear Regression
- Logistic Regression
- Decision Trees
- Neural Networks
- Support Vector Machines
- Evaluating Predictive Models

Main reference:

Baesens, B. (ed) (2012) Data collection, sampling, and preprocessing, in analytics in a big data world: The essential guide to data science and its applications. John Wiley & Sons.

6. Course Contents [8]

Lecture 7: Descriptive Analytics

- Association Rules
- Sequence Rules
- Segmentation

Main reference:

Baesens, B. (ed) (2012) Data collection, sampling, and preprocessing, in analytics in a big data world: The essential guide to data science and its applications. John Wiley & Sons.

6. Course Contents [9]

Lecture 8: Survival Analysis

- Survival Analysis Measurements
- Kaplan Meier Analysis
- Parametric Survival Analysis
- Proportional Hazards Regression
- Evaluating Survival Analysis Models

Main reference:

Baesens, B. (ed) (2012) Data collection, sampling, and preprocessing, in analytics in a big data world: The essential guide to data science and its applications. John Wiley & Sons.

6. Course Contents [10]

Lecture 9: Social Network Analytics

- Social Network Definitions
- Social Network Metrics
- Social Network Learning
- Relational Neighbor Classifier
- Probabilistic Relational Neighbor Classifier
- Relational Logistic Regression
- Collective Inferencing
- Egonets
- Bigraphs

Main reference:

Baesens, B. (ed) (2012) Data collection, sampling, and preprocessing, in analytics in a big data world: The essential guide to data science and its applications. John Wiley & Sons.

6. Course Contents [11]

Lecture 10: Example Applications

- Credit Risk Modeling
- Fraud Detection
- Net Lift Response Modeling
- Churn Prediction
- Recommender Systems
- Web Analytics
- Social Media Analytics
- Business Process Analytics

Main reference:

Baesens, B. (ed) (2012) Data collection, sampling, and preprocessing, in analytics in a big data world: The essential guide to data science and its applications. John Wiley & Sons.

7. My Timetable

Dr. Raini Hassan

Head (Computer Science)

Department: Computer Science, Level 4, KICT Building

Email: hrai@iium.edu.my | hraini@gmail.com Phone no: 03 6196 5650

SEMESTER II 2018/2019

DAY / TIME	8 - 9	9 - 10	10 - 11	11 - 12	12 - 1	1 - 2	2 - 3	3 - 4	4 - 5
MON						B			
TUE			CSC 3305 Section 1		CSC 3303 Section 1	R			
WED						E			
THUR			CSC 3305 Section 1		CSC 3303 Section 1	A			
FRI						K			

CSC 3303 – Big Data Analytics | CSC 3305 – Data Science

NOTE: I AM AVAILABLE AT THE DEPARTMENT OF COMPUTER SCIENCE.

Please email for course consultation.

8. Course Timetable [1]

February 2019

	Mon	Tue	Wed	Thu	Fri	Sat	Sun
					1	3	2
WEEK 1	4	5 Public Holiday: Chinese New Year	6 Public Holiday: Chinese New Year	7 Welcome to the Big Data Analytics class!	8	9	10
WEEK 2	11	12 Lecture 1	13	14 Group Assignment 1	15	16	17
WEEK 3	18	19 Lecture 2	20	21 Lecture 2 Group Project Proposal	22	23	24
WEEK 4	25	26 Lecture 3	27	28 Lecture 3			

Lecture 1: Introduction to Big Data Analytics
Lecture 2: Collection, Sampling and Pre-Processing
Lecture 3: Data Analytics Lifecycle (DAL)

Coursework:
Group Assignment 1 (10%): What Happens when the Data Becomes Bigger?
Group Project Proposal (5%)

8. Course Timetable [2]

March 2019

	Mon	Tue	Wed	Thu	Fri	Sat	Sun
WEEK 4					1	2	3
WEEK 5	4	5 Quiz 1	6	7 Lecture 4	8	9	10
WEEK 6	11	12 Lecture 4	13	14 Lecture 5	15	16	17
WEEK 7	18	19 Test 1	20	21 BREAK	22 BREAK	23 BREAK	24 BREAK
WEEK 8	25	26 Lecture 5	27	28 Lecture 6	29	30	31

Lecture 4: Introduction to Parallel Processing/Programming
Lecture 5: Hadoop, HDFS & MapReduce
Lecture 6: Predictive Analytics

Coursework:
Quiz 1 (10%): All lectures covered so far
Test 1 (20%): All lectures covered so far

8. Course Timetable [3]

	Mon	Tue	Wed	Thu	Fri	Sat	Sun
	April 2019						
WEEK 9	1	2 Lecture 6	3	4 Group Assignment 2	5	6	7
WEEK 10	8	9 Lecture 7	10	11 Lecture 8	12	13	14
WEEK 11	15	16 Quiz 2	17	18 Lecture 8	19	20	21
WEEK 12	22	23 Lecture 9	24	25 Lecture 9	26	27	28
WEEK 13	29 [IOW]	30 [IOW] Lecture 10					

Lecture 6: Predictive Analytics
Lecture 7: Descriptive Analytics
Lecture 8: Survival Analysis
Lecture 9: Social Network Analytics
Lecture 10: Example Applications

Coursework:

Group Assignment 1 (10%): All lectures covered so far
Quiz 2 (10%): All lectures covered so far

8. Course Timetable [4]

May 2019

	Mon	Tue	Wed	Thu	Fri	Sat	Sun
WEEK 13			1 [IOW] Public Holiday: Labor Day	2 [IOW] Lecture 10	3 [IOW] Group Project Report	4	5
WEEK 14	6	7 Group Project Presentation	8	9 Test 2	10	11	12
	13	14	15	16	17	18	19 Hari Wesak
	20 Public Holiday: Cuti Hari Wesak	21	22 Public Holiday: Nuzul Al-Quran	23	24	25	26
	27	28	29	30	31		

Coursework:

Group Project Report (10%)

Group Project Presentation (5%)

Test 1 (20%): All lectures

9. Additional Notes [1]

In the coursework, especially the group project, you are required to take initiatives to find the alternative ways (other than those covered in the classes) to solve your research problem. Most probably, you will be using the Google a lot. Please be extra careful on the sources and do acknowledge their works (as references in your report). The common mistakes/offences are:

Cheating – what is cheating?

- Cheating is an act of using dishonest means/ways or finding an easy way out for the purpose of getting rewards or advantages to oneself.
- Example 1: sneaking at friend's midterm paper when the lecturer is not looking.
- Example 2: asks friend for the answers when the lecturer is occupied.

9. Additional Notes [2]

Plagiarism – what is plagiarism?

- Plagiarism is an act of stealing someone's ideas, and it is also a form of cheating, which is morally and ethically unacceptable.
- Example 1: copying friend's assignment.
- Example 2: copy-and-paste information from the webpages or books without rephrasing (paraphrasing).

Paraphrasing – what is paraphrasing?

- Expressing an idea, question, fact in an alternative way.
- Example:
  - Original sentence: Kuala Lumpur has many skyline buildings, such as the Petronas Twin Towers.
  - Paraphrase sentence: Petronas Twin Towers is one of the skyline buildings that Kuala Lumpur has.

9. Additional Notes [3]

- **DO NOT BE LATE**, come 5 minutes early or be on time.

- **NO MAKE-UP QUIZZES, ASSIGNMENTS OR TESTS**
  - ***Unless:*** you have the medical certificate (MC), however only 2 MCs acceptable per semester, and must be submitted to me within 2 weeks. Else your MC is no good.
  - Time slip is not acceptable. There is a reason as to why the Doctor is giving you the time slip. It means you are not sick enough to skip the class.
  - **No make-up if you miss 2 classes without valid excuses (including the first week of the semester).**

9. Additional Notes [4]

- **NO LATE SUBMISSION OF ASSIGNMENTS/REPORT**
  - Beware of marks reduction.
- **DISCIPLINE: DON'T MISS ANY CLASSES**
  - If you miss 3 to 4 classes without valid excuses, a warning letter will be issued.
  - If you miss the next classes and MIA "missing-in-action", barring will be applied to you.

9. Additional Notes [5]

STRICT REMINDER

- **I WILL NOT provide any additional coursework to increase/fix the CAM (Carry Assessment Mark or Carry Forward of Marks).**
- **It is up to the student, as his/her responsibility to set his/her CAM.**
- **The basic but very important guidelines are:**
  1. **IF Efforts = Poor level, THEN Marks = POOR**
  2. **IF Efforts = Medium level, THEN Marks = MEDIUM**
  3. **IF Efforts = High level, THEN Marks = HIGH**
- **It is as simple as that and universally applied in every person's life on this earth; YOU ARE NOT AN EXCEPTION.**

Dua After Studying

Dua After Studying

اللَّهُمَّ إِنِّي أَسْتَوْدِعُكَ مَا قَرأتُ وَمَا حَفَظْتُ، فَرُضْهُ عَليّ عِنْدَ حَاجَتِي إِلَيهِ، إِنّكَ عَلى مَا تَشَاءُ قَدِيرُ وَأَنْتَ حَسْبِي وَنِعْمَ الوَكِيل

Allahumma inni astaodeeka ma qara'tu wama hafaz-tu. Farudduhu 'allaya inda hajati elahi. Innaka 'ala ma-tasha'-u qadeer wa anta hasbeeya wa na'mal wakeel.

Oh Allah! I entrust you with what I have read and I have studied. Oh Allah! Bring it back to me when I am in need of it. Oh Allah! You do whatever you wish, you are my availer and protector and the best of aid.