Computational Thinking and Big Data

COURSE DETAILS

No prerequisites required.

COURSE DATES

This is a self-paced course, so you can learn when it suits you.

Finish Date:
1st March 2022 0:00 AM UTC

TIME COMMITMENT

Between 8 to 10 hours per section.

ASSESSMENTS & CERTIFICATION

To qualify for a certificate, all questions, practical activities and assignments must be completed. edX will only issue certificates to participants that have chosen the 'Verified Track' and complete the course with a grade of 60% or higher. When your certificate is available, you will be notified in your edX dashboard.

GRADING SCHEME

Pass (60% or higher)
Fail (under 60%)

Related courses as part of the AdelaideX Big Data MicroMasters Program:

– Programming for Data Science

– Big Data Fundamentals

– Big Data Analysis

– Big Data Capstone Project

Find out more

Course syllabus

COURSE OVERVIEW

Computational thinking is an invaluable skill that can be used across every industry, as it allows you to formulate a problem and express a solution in such a way that a computer can effectively carry it out.

In this course, you will learn core computational thinking concepts including decomposition, pattern recognition, abstraction, and algorithmic thinking, and how to apply these in data science.

You will learn about data representation and analysis and the processes of cleaning, presenting, and visualizing data. You will develop skills in data-driven problem design and algorithms for big data.

This course will also explain mathematical representations, probabilistic and statistical models, dimension reduction and Bayesian models.

You will use tools such as R, Java and data processing libraries in associated language environments.

WHAT YOU WILL LEARN

» Understand and apply advanced core computational thinking concepts to large-scale data sets

» Use industry-level tools for data preparation and visualisation, such as R and Java

» Apply methods for data preparation to large data sets

» Understand mathematical and statistical techniques for extracting information from large data sets and illuminating relationships between data sets.

Section 1: Data in R

SECTION 1 LEARNING OBJECTIVES

» Identify the components of RStudio

» Examine data frames in R

» Identify the subjects and types of variables in R

» Summarise and visualise univariate data, including histograms and box plots.

SECTION 1 ASSESSMENT REQUIREMENTS

» Quiz questions and activities.

Section 2: Visualising relationships

SECTION 2 LEARNING OBJECTIVES

- Produce plots in ggplot2 in R to illustrate the relationship between pairs of variables
- Understand which type of plot to use for different variables
- Identify methods to deal with large datasets.

SECTION 2 ASSESSMENT REQUIREMENTS

- Quiz questions and activites.

Section 3: Manipulating and joining data

SECTION 3 LEARNING OBJECTIVES

- Organise different data types, including strings, dates and times
- Filter subjects in a data frame, select individual variables, group data by variables and calculate summary statistics
- Join separate dataframes into a single dataframe
- Learn how to implement these methods in mapReduce.

SECTION 3 ASSESSMENT REQUIREMENTS

- Quiz questions and activity.

Section 4: Transforming data and dimension reduction

SECTION 4 LEARNING OBJECTIVES

- Transform data so that it is more appropriate for modelling
- Use various methods to transform variables, including q-q plots and Box-Cox transformation, so that they are distributed normally
- Reduce the number of variables using PCA
- Learn how to implement these techniques in modelling data with linear models.

SECTION 4 ASSESSMENT REQUIREMENTS

- Quiz questions and activities.

Section 5: Summarising data

SECTION 5 LEARNING OBJECTIVES

- Estimate model parameters, both point and interval estimates
- Differentiate between the statistical concepts, parameters and statistics
- Use statistical summaries to infer population characteristics
- Utilise strings
- Learn about k-mers in genomics and their relationship to perfect hash functions as an example of text manipulation.

SECTION 5 ASSESSMENT REQUIREMENTS

- Quiz questions, activity and assignment.

Section 6: Introduction to Java

SECTION 6 LEARNING OBJECTIVES

» Use complex data structures
» Implement your own data structures to organise data
» Explain the differences between classes and objects
» Motivate object-orientation.

SECTION 6 ASSESSMENT REQUIREMENTS

» Quiz questions and activity.

Section 7: Graphs

SECTION 7 LEARNING OBJECTIVES

» Encode directed and undirected graphs in different data structures, such as matrices and adjacency lists
» Execute basic algorithms, such as depth-first search and breadth-first search
» Write graphs to files and read from them.

SECTION 7 ASSESSMENT REQUIREMENTS

» Quiz questions and activity.

Section 8: Probability

SECTION 8 LEARNING OBJECTIVES

» Determine the probability of events occurring when the probability distribution is discrete
» How to approximate e^x.

SECTION 8 ASSESSMENT REQUIREMENTS

» • Quiz questions.

Section 9: Hashing

SECTION 9 LEARNING OBJECTIVES

» Apply hash functions on basic data structures in Java
» Implement your own hash functions and execute, these as well as built-in ones
» Differentiate good from bad hash functions based on the concept of collisions.

SECTION 9 ASSESSMENT REQUIREMENTS

» Quiz questions and activities.

Section 10: Bringing it all together

SECTION 10 LEARNING OBJECTIVES

- Understand the context of big data in programming
- Transform a problem description into a complete working solution using the skills and knowledge you've learned throughout the entire course.

SECTION 10 ASSESSMENT REQUIREMENTS

- Major assignments.

Part of your journey in achieving the AdelaideX Big Data MicroMasters credentialential

This course is the third in a series of 5 courses that make up the Big Data MicroMasters Program from AdelaideX - a series of crediteligible courses that provides you with a graduate-level foundation in data science.

The Big Data MicroMasters has been designed for you to develop both the technical and computational skills that are in high demand across a range of industries.

Earn the industry-recognised MicroMasters credential by completing and successfully earning a Verified Certificate in all 4 Big Data courses plus the Capstone Project course. Go to:

- https://www.edx.org/micromasters/adelaidex-big-data

for further information and links to all the related courses.

Take your credential to the next level with a Master of Data Science degree

Learners who successfully complete the Big Data MicroMasters program and successfully gain admission to the Master of Data Science (Applied) Online at the University of Adelaide will be granted 12 units of credit towards the Master's degree program.

The Big Data MicroMasters credential represents 25% of the Master of Data Science degree program, which requires 48 units of coursework to complete.

Discussion Forum Etiquette and Frequency

We expect you to follow the edX rules for online conduct at all times and keep your posts/responses positive on the discussion forums.

Post regularly, at least once per discussion activity and be sure to respond to your peers, as instructed.