Big Data Fundamentals

COURSE DETAILS

Candidates interested in pursuing the MicroMasters program in Big Data are advised to complete Programming for Data Science and Computational Thinking and Big Data before undertaking this course.

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

– Computational Thinking and Big Data

– Big Data Analysis

– Big Data Capstone Project

Find out more

Course syllabus

COURSE OVERVIEW

Organisations now have access to massive amounts of data and it's influencing the way they operate. They are realising in order to be successful they must leverage their data to make effective business decisions.

In this course, part of the Big Data MicroMasters program, you will learn how big data is driving organisational change and the key challenges organisations face when trying to analyse massive data sets.

You will learn fundamental techniques, such as data mining and stream processing. You will also learn how to design and implement PageRank algorithms using MapReduce, a programming paradigm that allows for massive scalability across hundreds or thousands of servers in a Hadoop cluster.

You will learn how big data has improved web search and how online advertising systems work.

By the end of this course, you will have a better understanding of the various applications of big data methods in industry and research.

WHAT YOU WILL LEARN

» Knowledge and application of MapReduce

» Understanding the rate of occurrences of events in big data

» How to design algorithms for stream processing and counting of frequent elements in Big Data

» Understand and design PageRank algorithms

» Understand underlying random walk algorithms.

Section 1: Web and social networks

SECTION 1 LEARNING OBJECTIVES

» Understand the four Vs of Big Data (Volume, Velocity, Veracity and Variety)

» Build models for data

» Understand the occurrence of rare events in random data.

SECTION 1 ASSESSMENT REQUIREMENTS

» Quiz questions.

Section 2: Clustering big data

SECTION 2 LEARNING OBJECTIVES

> » Understand characteristics of the web and social networks
> » Model social networks
> » Apply algorithms for community detection in networks.

SECTION 2 ASSESSMENT REQUIREMENTS

> » Quiz questions and activites.

Section 3: Clustering big data

SECTION 3 LEARNING OBJECTIVES

> » Clustering social networks
> » Apply hierarchical clustering
> » Apply k-means clustering

SECTION 3 ASSESSMENT REQUIREMENTS

> » Quiz questions and activity.

Section 4: Google web search

SECTION 4 LEARNING OBJECTIVES

> » Understand the concept of PageRank
> » •mplement the basic PageRank algorithm for strongly connected graphs
> » Implement PageRank with taxation for graphs that are not strongly connected

SECTION 4 ASSESSMENT REQUIREMENTS

> » Quiz questions and activities.

Section 5: Parallel and distributed computing using MapReduce

SECTION 5 LEARNING OBJECTIVES

> » Understand the architecture for massive distributed and parallel computing
> » Apply MapReduce using Hadoop
> » Compute PageRank using MapReduce

SECTION 5 ASSESSMENT REQUIREMENTS

> » Quiz question, activities and major assignment 1.

Section 6: Computing similar documents in big data

SECTION 6 LEARNING OBJECTIVES

> » Measure importance of words in a collection of documents
> » Measure similarity of sets and documents
> » Apply local sensitivity hashing to compute similar documents.

SECTION 6 ASSESSMENT REQUIREMENTS

> » Quiz questions and activity.

Section 7: Products frequently bought together in stores

SECTION 7 LEARNING OBJECTIVES

» Understand the importance of frequent item sets

» Design association rules

» Implement the A-priori algorithm.

SECTION 7 ASSESSMENT REQUIREMENTS

» Quiz questions and activity.

Section 8: Movie and music recommendations

SECTION 8 LEARNING OBJECTIVES

» Understand the differences of recommendation systems

» Design content-based recommendation systems

» Design collaborative filtering recommendation systems.

SECTION 8 ASSESSMENT REQUIREMENTS

» Quiz questions and activity.

Section 9: Google Ads™ System

SECTION 9 LEARNING OBJECTIVES

» Understand the Google Ads System

» Analyse online algorithms in terms of competitive ratio

» Use online matching to solve the Google Ads problem

SECTION 9 ASSESSMENT REQUIREMENTS

» Quiz questions and activity.

Section 10: Mining rapidly arriving data streams

SECTION 10 LEARNING OBJECTIVES

» Understand types of queries for data streams

» Analyse sampling methods for data streams

» Count distinct elements in data streams

» Filter data streams.

SECTION 10 ASSESSMENT REQUIREMENTS

» Quiz questions, activity and major assignment 2.

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.