Course Title

Mastering Big Data Analytics with Hadoop

Course Description

The ability to collect huge amounts of data has led to the exponential growth in the size of data repositories. Companies such as Google, Facebook, Amazon, Ebay, Microsoft, IBM, etc. all perform in depth analysis on these huge repositories to gain valuable actionable insights. The key Big Data technology used by all these companies is the Hadoop Big Data processing platform. In depth knowledge in Hadoop is currently one of the most sought after skills in Silicon Valley. Quote from a Facebook engineer, “Hadoop is used in every Facebook product and in a variety of ways”.

In this course you will learn to use the latest Big Data technologies from the Hadoop Ecosystem to perform in depth analytics over Big Data. Hadoop has undergone a massive change in the last 3 years. In the past Hadoop was only useful for batch processing of Big Data. However, in recent years, many new tools have emerged from the Hadoop ecosystem that enable interactive analysis and machine learning over huge data sets. In this course we will dive deep into these latest and most exciting technologies while at the same time give you a solid grounding in the fundamentals of how Hadoop works. Currently the hottest and most powerful programming framework within the Hadoop ecosystem is the Spark framework. People in the Berkley University developed this framework. Spark is up to 100x faster than the existing MapReduce framework and can be used to do every type of analytics including, SQL queries, machine learning, graph analytics, stream processing, etc. This is the only Big Data course in Australia to teach Spark.

This course includes both lectures and hands on programming. There is a strong emphasis on gaining hands on experience. Therefore, the majority of the course will be spent on programming solutions to Big Data problems using the latest tools.
Outline of 3-day course

Day 1

1. Introduction to Hadoop
2. Lecture on the MapReduce framework including HDFS
3. Lecture on how to program MapReduce using Java
4. Basic MapReduce programming exercises using the Cloudera Virtual Machine.
5. More advanced MapReduce programming exercises using the Cloudera Virtual Machine

Day 2

1. Introduction to Hive (SQL on top of MapReduce)
2. Programming exercises in Hive using Cloudera Virtual Machine
3. Brief overview of the entire Hadoop ecosystem
4. Introduction to Spark and Scala
5. Programming exercises in Scala

Day 3

1. Lecture on more advanced concepts in Spark
2. Lecture on how to program in Spark
3. Programming exercises in Spark on Cloudera Virtual Machine
4. Programming exercises in Spark on Amazon's Elastic Map Reduce Service
Assumed Knowledge

- No prior knowledge of Hadoop is assumed.
- The student should have at least basic Java programming experience.
- The student should have basic SQL Programming experience.

Time and Dates

Dates: Wednesday 25\textsuperscript{th} of June to Friday 27\textsuperscript{th} of June
Time: 10am – 5pm each day.

Cost

$2200 for non La Trobe University staff (inclusive of GST)
$1760 for La Trobe University staff

Location

Beth Gleeson Building ground floor
Department of Computer Science and Computer Engineering
La Trobe University
Kingsbury Drive
Bundoora

Contact Person
Name: Dr Zhen He
Phone: 94793036
Email address: z.he@latrobe.edu.au

Registration

To register for the course please email Zhen He via the email address above.
Please register early. Registrations closes on Friday 20\textsuperscript{th} of June.