# Spark巨量資料處理與開發實務


授課講師 邱士軍

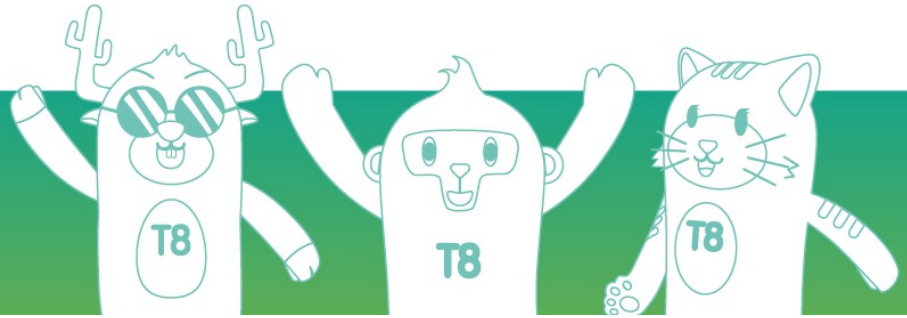
教材編寫 邱士軍



緯育 TibaMe

即學・即戰・即就業

https://www.tibame.com/

緯育 TibaMe

# 課程大綱

- Module 1. Big data eco-system Introduction
- Module 2. Spark/Hadoop development environment Setup
- Module 3. Demo - Multi-machine deployment for Spark and Hadoop
- Module 4. Spark core features overview
- Module 5. Demo - Dive-in with WordCount calculation
- Module 6. SparkContext, RDD, Transformations and Actions overview
- Module 7. Spark architecture and execution flow introduction
- Module 8. Common RDD Transformation and Actions I
- Module 9. Demo - Web logs analysis for detecting traffic fraud and identify locations
- Module 10. Common RDD Transformation and Actions 2

緯育 TibaMe

# 課程大綱

- Module 11. Demo - Movie data analysis for movie similarity for movie recommendation
- Module 12. RDD Persistence and Accumulator and Broadcast variables
- Module 13. Spark SQL overview and programming model, I
- Module 14. Spark SQL overview and programming model II
- Module 15. Creating DataFrame from structural semi-structural files - JSON and CSV files
- Module 16. Real-world file format for storing big data
- Module 17. Creating DataFrame from RDD, Drive-local list and RDB tables
- Module 18. DataFrame analysis using SQL and create UDFs
- Module 19. Demo - Creating UDFs and web logs analysis demo
- Module 20. DataFrame analysis using Transformation API I

緯育 TibaMe

# 課程大綱

- Module 21. DataFrame analysis using Transformation API II
- Module 22. Demo - Crime and Finance data analysis
- Module 23. Spark Streaming overview and programming model
- Module 24. StreamingContext, DStream, Streaming program console introduction and performance monitoring
- Module 25. File stream introduction and demo
- Module 26. DStream transformations and window methods
- Module 27. Receiving data from Kafka
- Module 28. Demo - Real-time sentiment analysis with a pre-trained model from Kafka traffic
- Module 29. DStream output actions
- Module 30. Demo - Real-time EC logs analysis