

# INTRODUCTION TO SPARK WITH SCALA

## INTRODUCTION

# Course Overview

2

- Syllabus
- Grading
- Resources
- Getting to know each other

# About Me

- Hien Luu
  - ▣ Engineering Manager @ LinkedIn
  - ▣ Big Data products, applications and infrastructure
- Been teaching for > 10 years @ UCSC
  - ▣ Apache Spark
  - ▣ Apache Hadoop
  - ▣ Spring Framework
  - ▣ Java Comprehensive
- Email: [hluu@yahoo.com](mailto:hluu@yahoo.com)

# Course Overview

4

## □ Syllabus

- ▣ Spark architecture & execution model
- ▣ Introduction to Scala
- ▣ Spark DataFrame Programming Model
- ▣ Spark SQL, UDF & Advanced Functions
- ▣ Spark Streaming & Structured Streaming
- ▣ Machine Learning & Spark MLlib

# Course Overview

5

## □ Grading

### ▣ 4 Assignments

- 1 Scala assignment
- 1 Read RDD paper and answer questions
- 2 programming assignments

## □ 10 Sessions

### ▣ Tuesday 04/07 – 06/09

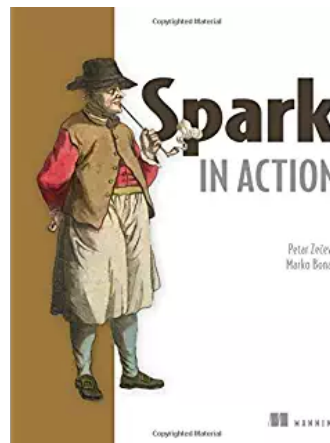
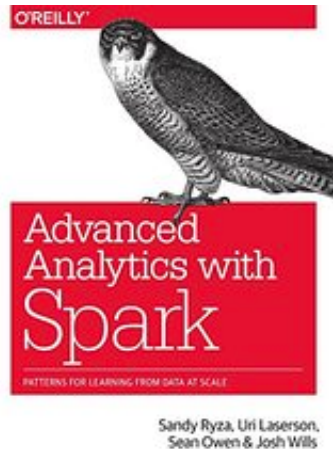
## □ Class Hours

- ▣ 6:30pm – 9pm
- ▣ 1 break around 8pm

# Course Overview

6

## □ Resources - Books



# Resources

7

- Google site for Spark class
  - ▣ <https://sites.google.com/site/continuelearning/Home/introduction-to-spark-with-scala>
- Databricks Community Edition - databricks.com/ce
  - ▣ Everyone should sign up for an account

# Getting to know each other

8

- ☐ Name
- ☐ What do you do? And where?
- ☐ Software development background?
- ☐ Why are you taking this class?
- ☐ What's your motivation?

Please keep it short