

# Course Introduction

## Cloud Computing and Big Data (CLO)

Oxford University  
Software Engineering  
Programme  
Sep 2016



# Introduction

- Aims
- Pre-requisites
- Contents
- Objectives
- Resources
- Rules of Engagement
- Introductions



# Aims

- Understanding of Principles of Cloud Computing and Big Data
- Theoretical background and origins
- Practical experience of different technologies
- Architecture and Design
- Wider context



# Pre-requisites

## Covered by the Pre-Study Guide

- **Command line** tooling and Unix commands
- Some **Python programming** and **text editors**
- **SQL** and data manipulation
- **Understanding** of networking, servers and distributed computing



# Format

- A mixture of lectures and practical labs
- Lectures aim to provide the wider context and background
  - Independent of specific technologies
- Labs are based on specific technologies
  - Designed to demonstrate the principles



# Lab model

- Local Virtual Machine
  - Ubuntu
  - Pre-installed big data software
    - E.g. Apache Hadoop and Spark, Docker, etc
- Amazon Web Services
  - Virtual machines in the cloud



# Contents

- Overview and Introduction
- Cloud Computing
  - Introduction and Case Studies
  - Cloud Computing Theory and Background
  - Containers and Docker
- Big Data
  - Introduction and Case Studies
  - Map Reduce and Hadoop
  - Apache Spark and in-memory big data
  - Realtime
  - Visualisation
  - NoSQL
  - Cassandra



# Practicals

- Using Cloud Services
- Elastic scaling
- Hadoop and Map Reduce
- Spark, SparkSQL
- Cassandra and NoSQL
- Spark and Cassandra together
- Realtime big data
- Containers
- Visualisation





# Specific Objectives

- Understand the principles of cloud computing
  - Theory of scalability
  - Including scalability and deployment
  - IaaS frameworks, PaaS, containers
- Understand Big Data approaches, technologies and techniques
  - Theoretical background and approaches
  - Including Map Reduce, NoSQL, Realtime
- Be able to design and implement scalable cloud and big data systems
- Understand and implement effective Open Source systems on Amazon EC2



# Improve your CV?



Leverage the NoSQL boom

# Beyond the scope of this course

- Detailed Data Science techniques
- Implementing a private cloud
  - Although we will look at technologies for private cloud
- Understanding all of Hadoop, Spark, Mesos, CoreOS, etc



# Rules of Engagement

- ***Ask questions as we go along***
  - We will “park” any that are better answered later
  - Don’t wait till the end to ask or raise concerns
  - If you don’t ask we can’t help you



# There ~~might~~ will be bugs!



- Please help out:
  - Please create new issues on the Github repository
  - <https://github.com/pzfreo/ox-clo/issues/new>

# Paul Fremantle

- CTO and Co-Founder of WSO2
  - Currently on study leave
- Previously Senior Technical Staff Member, IBM WebSphere architecture
- VP, Apache Synapse and Member of ASF
- BA in Maths and Philosophy
- MSc in Computation (1995)
- Teaches SOA module
- PhD Research Student



# You?



© Paul Fremantle 2015. This work is licensed under a Creative Commons  
Attribution-NonCommercial-ShareAlike 4.0 International License  
See <http://creativecommons.org/licenses/by-nc-sa/4.0/>

# Approximate Schedule

Monday	Tuesday	Wednesday	Thursday	Friday
Overall Introductions	Introduction to Big Data and case studies	Spark and SQL	Visualisation	Overview and Recap
First Cloud lab exercise	Hadoop Lab 1	SparkSQL Lab	Visualisation Lab	Presentation
Cloud Overview and case studies	Hadoop details, Map-Reduce	Storage and NoSQL	Containers	Group Exercise
Elastic Cloud Lab	Hadoop Extras	Cassandra Lab	Docker Lab	Final Thoughts and Assignment
Cloud Theory Platform-as-a-Service, scaling	Intro to Spark	Cassandra details	Realtime Big Data	
Further Cloud Lab		Cassandra Lab2	Realtime Lab	





# Let's get started



© Paul Fremantle 2015. This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License  
See <http://creativecommons.org/licenses/by-nc-sa/4.0/>