Course Introduction

Big Data Engineering in the Cloud

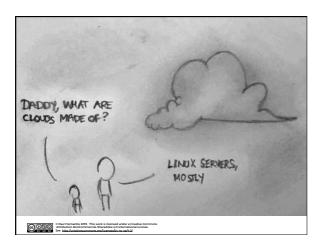
Dec 2017

© 19 Paul Fremantie 2015. This work is licensed under a Creative Common Attribution-NonCommercial-ShareAlike 4.0 International License

Introduction

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

© Paul Fremantie 2015. This work is licensed under a Creative Commo Attribution-NonCommercial-ShareAlike 4.0 international License See hits firest incommons conficences by re. sal 4.0/



Aims

- Understanding principles of Big Data
- Theoretical background and origins
- Practical experience of modern big data processing systems technologies
- · Architecture and design
- Wider context

© Paul Fremantie 2015. This work is licensed under a Creative Common Attribution-NonCommercial ShareAlice 4.0 International License See Intrins Districts Associations (NonCommons conficiences NonCommons Conficiences Non

Pre-requisites

Covered by the Pre-Study Guide

- Command line tooling and Unix commands
- Some Python programming and text editors

© Paul Fremantie 2015. This work is licensed under a Creative Common Attribution-NonCommercial-ShareAlike 4.0 International License See http://www.noncommons.com/licenses/bv-nc-as/4.0/

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

© Paul Fernantile 2015. This work is licensed under a Creative Common Attribution-NonCommercial ShareAlice 4.0 International License See http://creativecommons.com/licenses.bu-c-saf4.0/

Lab model

- · Local Virtual Machine
 - Ubuntu
 - Pre-installed big data software
 - E.g. Apache Spark, Cassandra, Python
- Amazon Web Services
 - Virtual machines in the cloud

© DOO Paul Fermantie 2015. This work is licensed under a Creative Common Attribution. NonCommencial ShareAlite 4.0 International License See Into Exercises See Into Exercises See Into Exercises See Into Exercises See Into

Contents

- Big Data motivation and overview
- Using Python for Data Analysis
- Map Reduce and Directed Acyclic Graphs
- · Apache Spark
- Spark and SQL
- Theory of scaling
- Running Spark on Amazon
- Introduction to NoSQL databases
- Introduction to Machine Learning

© Paul Fremantle 2015. This work is licensed under a Creative Common Attribution-NocCommercial ShareAlike 4.0 International License See Not 6.5 See Into 5.5 See Into 6.5 See

Practicals

- Python Data Analysis
- Spark, SparkSQL
- Spark on Amazon
- Cassandra and NoSQL
- Machine Learning libraries
- Visualisation

© Paul Fremantie 2015. This work is licensed under a Creative Common Attribution-NonCommercial ShareAlike 4.0 International License See http://creativecommons.com/licensesby-nc-sa/4.0/

Improve your CV? ROYOU HAVE NO SOLUTION AND THE PROPERTY OF THE PROPERTY OF

Beyond the scope of this course

- Detailed Data Science techniques
- Understanding **all** of Hadoop, Spark, HDFS, Machine Learning

DOO
 Paul Fremantle 2015. This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License See hits Offerathroomers could be consession or safe Offeration.

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

© Paul Fremantie 2015. This work is licensed under a Creative Common Attribution-NonCommercial ShareAlike 4.0 International License See http://creativecommons.com/licenses/bu-nc-sa/4.0/

There might will be bugs!



- · Please help out:
 - Please create new issues on the Cithub repository
 - https://github.com/pzfreo/big/issues/new



Paul Fremantle

- CTO and Co-Founder of WSO2
- Previously Senior Technical Staff Member, IBM WebSphere architecture
- Visiting Lecturer, Oxford University
- VP, Apache Synapse and Member of Apache
- PhD in Computing (2017)
 loT security and privacy





David Johnson

- · Senior Researcher, e-Science, Oxford University
- Founding Member, Data Science Institute, Imperial
- PhD, Reading University, 2010
- Awarded University of Oxford Teaching Award, 2016



| | - |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---|
| V2 | |
| You? | |
| | |
| | |
| | |
| | |
| | |
| The function 255. This work is bossed under a Continuous The Community of the Comm | |
| | |
| | |
| | |
| Approximate Schedule | |
| Weds Friday Introduction to Machine | |
| Overview and Motivation Data Analysis with Python Realtime systems | |
| and Pandas – Architecting big data – Map Reduce systems – Apache Spark – Completion of labs | |
| Thursday SQL Theoretical background on | |
| scaling systems – Scaling Spark on AWS | |
| Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualisation Visualis | |
| © Q Q Q Part Trement 2015. The var has known and under a Creative Common definition of the Common definition and Common definition | |
| | |
| | 1 |
| Let's get started | - |
| | |
| The state of the s | |
| | |
| | |