Fundamentals of Big Data

Outline

* Concepts around BD, DS, ML, AI…  
  Introduction to Data Science
  + Roles, Skills, Methods
  + Asking and Answering Insight Questions
  + Science in Business
* Introduction to Big Data
  + Data Models
  + NoSQL
  + CAP Theorem
  + Tools
* Python (Refresher) for Data Science
* Data Analytics & Statistical Inference
  + Machine Learning Methodology…
  + Workflow/Pipeline project
* Big Data with Neo4j
  + GraphDB
* Big Data with Mongo
  + Document Database (Key-Value)
* Big Data with Hadoop
  + Schemaless storage (file system)
  + Big Data with Spark
    - Distributed computations across hadoop

Admin:

8.30 to (3-3.30)   
Breaks finishing at 10am, 12.30pm, 1.30pm  
[michael.burgess@qa.com](mailto:michael.burgess@qa.com)

Process:

Advise:

* If not speak, please mute

Michael Burgess  
Principal for Data, Analytics and AI

Bio:  
QA for 5 years  
IT Contractor for 4/5years – Fintech, Defense, Web, …  
Physics 7yrs  
Discussions, Philosophy, Arguments…

* Prior experience with Python is required
  + Programming,..
* Experience working with data
  + Tools, Techniques (Statistics, Analysis)

Name:  
Role/Background:  
Relevant Experience:  
Expectations:

what are you looking to achieve/etc. by the end of the course…  
any particular interests, topics, you’d like to see

Hobbies/Interests: …

**Tobias**Computational fluid Dynamics, westing house,   
Modelling & Simulations, Condition Monitoring  
Technical Manager around fluid modelling  
General View, implementation possibilities (AI, ML, BD,…)  
Interests: Cross-country skiing

**Paul**Westing House 12yr, Nuclear 19yr, Physicist , Data analysis  
Model development, programming, data analysis  
Big Data (measurement) data sets & calculation sets  
More fortran, perl, -> Python  
How does big data relate to this field?  
Current problem: many types & disparate kinds, volume  
Single experience has millions of data points  
Reading, Films -> Children, DIY

**Jonathan**Masters student in nuclear engineering (at westing house)  
experience with various programming languages (at university)  
University modules on data analysis (monte carlo, variance reduction)  
introduction to overall topic  
hobbies: watching movies

**Gustav**Performance Modelling, Westing House 8yr  
Responsible for FORTRAN code, building models, validating against data  
Developing methodology for use of statistics (Markov, Gaussian Process)  
Test Reactor in Norway – big data solution around storing data  
Lots of data associated with measurements, collection problem  
Relational schema too constrained to handle flat-text file  
Tools around schemaless storage  
General question: how does big data fit into nuclear engineering?  
Hobbies: Statistics, Python, Parenting

**Erik**Thesis at Westing House, degree in engineering physics  
Some basic experience with python, programming in matlab/c  
basics & introduction  
news, choir, …

**Denise**Physicist, Nuclear Fuel development – 12yr  
High Performance Computing – some python for this…  
Python & Machine Learning, Data Analysis… (Keras,..)  
Tools around Data Handling/Engineering  
Video Games (xbox storymode)

**David**  
Westing House, 10 years  
General Analysis – Mechanical, Fluid,…  
Some python experience (java, c, ..)  
large amounts of data, specific accuracy in needs  
Hobbies: Parenting

**Camilla**   
Reactor Dynamics for 30 years – simulations & measurement   
Programming experience in various languages (eg., fortran, etc.)  
Orientation – concepts, ideas, techniques,…  
Philosophy (Epistemology of Computer Simulation)  
  
**Caj**   
Working for 30 years in automation at westing house  
Programming -> Automation Systems in Nuclear Plants  
Technical Project Manager  
Limited Python, online class  
Similar to ben – data from customers plants, analysis…   
Condition Monitoring  
Running, Bicycling, Mountaineer

**Ben**  
Working with software since 1992…   
Last 10 years with measurement and test systems  
Programming, Project Manager,…  
Technical Project Management  
Limited, some experience, python  
Figure out how to work with big data  
Measurement Data -> Analysis of Measurement Data  
Sea kayaking