College of Continuing Education & Professional Development

Big Data: Data Wrangling

Big Data: Data Wrangling

Digging into Big Data!

This three-day bootcamp, we will cover aspects common to all Big Data investigations, including: defining Big Data, surveying tools and techniques for processing Big Data, and visualizing selected aspects of Big Data. The focus of the camp will be primarily on velocity and variety.

The participants will use tweets from Twitter as a stand-in for velocity, and use the text from the tweets to provide variety. Selected tweets will be downloaded in real-time, stored in a local database, and then analyzed in near real-time. Ideas, concepts, and experiences gained from intensive classroom training and programming projects will be generalized to enterprise level systems to solve problems at scale.

The emphasis of the camp is to understand what is Big Data beyond the marketing hype of the 3Vs of volume, variety, and velocity, The bootcamp gives participants the opportunity to "cleanse" or "normalize" Big Data using Python and R, and construct Big Data processes to conduct sentiment analysis on real-time data.

This bootcamp is being offered with a non-credit or university credit option. The instructor will provide additional information to those who choose the credit-based option during registration.

About the Program:

Logistics

Dates: 3 days, September 16 - 18, 2016

Day of the week: Friday - Sunday

Times: 9:00 AM - 5:00 PM

Location:
Old Dominion University's Tricities Center
1070 University Blvd
Rooms 1230 A and B
Portsmouth, VA 23703

Contact hours: 24 (non-credit option) 30 (credit option)

Bootcamp Registration Pricing Options

Non-Credit Registration: $499.00

Credit Registration: CS 395, $659.00 + $50 application fee

Registration Deadline: September 11, 2016 **EXTENDED!**

Course Topics:

Defining big data, retrieving data using HTTP based application program interfaces (APIs), exploratory data analysis (EDA), storing (SQL vs. NoSL), using Python, and R.

Target Audience:

College and University Students, Military, Software Managers, Software Developers, Engineers, Developers, Architects, Networking specialists, Managers, Executives, Professional Services, Data Analyst, QA, Performance Engineers, Data Warehouse Professional, Sales, Pre Sales, Technical Marketing, PM, Teaching Staff, Delivery Manager

Benefits:

Upon completion of this course students will be able define and characterize Big Data, analyze, design, and implement real-time programs that retrieve data using HTTP based application program interfaces (APIs), store data in an SQL database, analyze the data to detect sentiment and trends.

Daily Schedule:

Day One:

A presentation with question and answer period covering: defining big data and big data processing overview. Topics covered include terms, ideas, and vocabulary, tools, exploratory data analysis (EDA), storing (SQL vs. NoSQL), and more.

Day Two:

Processing of "interesting" data with Python

Day Three:

Processing of "interesting" data with R

Learning Methods:

1. Lectures
2. Hands-on Exercises
3. Readings

Optional Materials:

Materials are not included in the cost of the course.

Data Wrangling with Python, by Kazil and Jarmul (ISBN: 1491948817)

o Click here to order on Amazon (http://www.amazon.com/Data-Wrangling-Python-Tools-Easier/dp/1491948817)

Learning Objectives:

Participants will be able to:

1. Enumerate the characteristics that qualify a problem as a Big Data problem requiring Big Data tools and techniques.
2. Identify issues associated with handling and manipulating Big Data
3. Identify different types of databases (SQL and Non-SQL) and what kinds of data and analysis each is optimized for
4. Apply elementary statistics to describe Big Data as part of exploratory data analysis (EDA)
5. Collect real-time data via HTTP based APIs using different languages
6. Store real-time data into different databases
7. Analyze real-time data with different languages
8. Visualize real-time data

Refunds/Cancellations

To withdraw from a course you must send a request in writing to rclaud@odu.edu (mailto:rclaud@odu.edu) seven (7) days prior to the start date of the course. Failure to attend a course does not constitute withdrawal. Course registration fees, less a $50 processing fee will be refunded via check per Old Dominion University policy (we are unable to return money to a debit or credit card).

There are no refunds once the class has begun.

Late withdrawals of six (6) days or less before the class begins, will result in the student being charged the $50 processing fee, as well as charges for books and/or other course material fees.

About the Faculty

Dr. Charles Cartledge has over 30 years of hands-on experience solving hard real-time Big Data processing. He has developed distributed processing systems integrating, consolidating, and visualizing data from disparate data sources where each bit of each byte is important. Dr. Cartledge has developed and taught Big Data and NoSQL Computer Science courses at the senior and Master's level. The courses covered the spectrum of Big Data tools, techniques, and languages including: Hadoop, Pig, MapReduce, HDFS, neo4J, HBse, MongoDB, and Redis.

MAKE YOUR MOVE!

Register Now! - Noncredit

Apply Now! - College Credit

Non-ODU Students

ODU Students & Alumni

*** To register and pay by check, please make your check payable to:**

Old Dominion University
1009 Rollins Hall
Norfolk, VA 23529

Contact Us:

For more information about this program, contact Dean Claud by e-mail at rclaud@odu.edu (mailto:rclaud@odu.edu) or call 757-683-4232.

Updated: 9/15/2016