Robusta Technology & Training

Trụ sở chính Tp. Hồ Chí Minh
97-99-101 Nguyễn Công Trứ, Quận 1, Tp. HCM
Tel: (+84) 8.54399188 | Hotline:(+84) 939 586 168
Website: www.Robusta.vn | Email: Learn@Robusta.vn

Chi nhánh Hà Nội
Tầng 5, số 17, Ngõ 167 Tây Sơn, Đống Đa, Hà Nội
Tel: (+84) 4.62930 889

Các phương pháp nâng cao trong Khoa học dữ liệu và Phân tích Dữ liệu lớn

I. Tổng quan về khoá học

Khóa học Các phương pháp nâng cao trong Data Science và Big Data Analytics cung cấp các kiến thức thực tế giúp người học có thể ngay lập tức tham gia vào các dự án Data Science và Big Data Analytics. Khóa học này giới thiệu về Big Data và Data Analytics Lifecycle nhằm giải quyết các yêu cầu liên quan đến Big Data trong việc kinh doanh. Theo đó, ngoài những kiến thức căn bản, khóa học còn cung cấp cho người học những phương pháp phân tích nâng cao, đề cập đến công nghệ và các công cụ phân tích dữ liệu lớn, bao gồm: MapReduce và Hadoop. Hệ thống Labs được cấp cho học viên giúp người học hiểu được về cách thức mà các phương pháp và công cụ được áp dụng vào trong thực tế kinh doanh. Môi trường của khóa học là môi trường "mở", dựa theo cách tiếp cận bán kỹ thuật. Kết thúc khóa học, học viên trải qua bài lab cuối khóa với mục tiêu giải quyết các vấn đề phân tích Big Data bằng cách áp dụng những kiến thức được giảng dạy trong toàn bộ nội dung chương trình Data Analytics Lifecycle. Sau khóa học, học viên đã tham gia khóa học có khả năng thi chứng chỉ Proven™ Professional Data Scientist Associate (EMCDSA).

II. Thời lượng khóa học: 40 giờ

III. Hình thức đào tạo: Giảng dạy trực tiếp tại lớp học.

IV. Mục tiêu khóa học

Sau khi kết thúc khóa học, học viên đạt được:

- Khai thác và sử dụng được các chức năng của MapReduce.
- Kết hợp nhuần nhuyễn giữa NoSQL databases và công cụ Hadoop Ecosystem cho việc phân tích các dữ liệu lớn và phi cấu trúc.
- Nâng cao kiến thức nghiệp vụ về Natural Language Processing (Xử lý ngôn ngữ tự nhiên), Social Network Analysis (phân tích mạng xã hội), và Data Visualization Concepts.
- Nắm bắt các phương pháp định lượng nâng cao, và áp dụng một trong số các giải pháp đó vào môi trường Hadoop.
- Ứng dụng các kỹ thuật nâng cao trên môi trường dữ liệu thực tế vào bài lab cuối khóa.

V. Đối tượng tham gia khóa học:

- Chuyên viên phân tích dữ liệu
- Có kiến thức cơ bản về Data Science và Big Data
- Có kiến thức cơ bản về lập trình và thống kê

VI. Nội dung chi tiết khóa học:

Nội dung của khóa học được biên soạn để đạt được các mục tiêu đề ra:

Module 1: MapReduce và Hadoop

- Lesson 1: MapReduce Framework
- Lesson 2: Apache Hadoop
- Lesson 3: Hadoop Distributed File System
- Lesson 4: YARN

Module 2: Hadoop Ecosystem và NoSQL

- Lesson 1: Hadoop Ecosystem
- Lesson 2: Pig
- Lesson 3: Hive
- Lesson 4: NoSQL - Not Only SQL
- Lesson 5: HBase
- Lesson 6: Spark

Module 3: Xử lý ngôn ngữ tự nhiên (Natural Language Processing – NPL)

- Lesson 1: Giới thiệu NLP
- Lesson 2: Text Preprocessing
- Lesson 3: TFIDF
- Lesson 4: Beyond Bag of Words
- Lesson 5: Language Modeling
- Lesson 6: POS Tagging and HMM
- Lesson 7: Sentiment Analysis và Topic Modeling

Module 4: Social Network Analysis

- Lesson 1: Giới thiệu SNA và Graph Theory
- Lesson 2: Most Important Nodes
- Lesson 3: Communities and Small World
- Lesson 4: Network Problems and SNA Tools

Module 5: Data Science Theory and Methods

- Lesson 1: Simulation
- Lesson 2: Random Forests
- Lesson 3: Multinomial Logistic Regression

Module 6: Data Visualization

- Lesson 1: Perception và Visualization
- Lesson 2: Visualization of Multivariate Data