

교육 과정 소개서.

Apache Spark을 이용한 빅데이터 분석 입문 CAMP



코스요약

코스명	Apache Spark을 이용한 빅데이터 분석 입문 CAMP
기간	2020.02.05 - 04.29 (4월 15일 휴강)
일정	매주 수요일 19:30 - 22:30
장소	패스트캠퍼스 강남강의장
준비물	개인 노트북(RAM 8GB 이상 권장), 필기구
담당자	02-568-9886 / help-ds@fastcampus.co.kr
수강료	1,800,000
상세페이지 url	fastcampus.co.kr/data_camp_spark

코스목표

데이터 분석의 전 과정을 단일 프로그래밍 환경에서 ALL-IN-ONE으로 끝낸다! 조사 시간은 반으로, 생산성은 두 배로! 더 쾌적한 당신의 분석 작업을 만들어 줄 SPARK를 만나세요.

코스정보

SPARK 앞에만 서면 작아지는 당신을 위해, 국내 최고의 SPARK 입문자용 코스를 준비했습니다. 데이터 수집- Spark 인메모리 데이터 처리(데이터 분산처리) - Spark SQL(데이터 전처리) -Spark MLib(데이터 분석) 순서에 맞추어 커리큘럼이 구성되어 있습니다. 수업 말미에는 여러분이 직접 음악 시스템을 구현해보는 실습 시간 또한 준비되어 있는 코스입니다.



코스특징

Spark 입문자에게 필요한 필수 개념 총망라

Spark의 핵심인 Spark Core은 기본, 현업에서 활용도가 높은 Spark SQL, Spark Streaming, Spark MLlib까지! 독학이 막연했던 리눅스와 인프라, Hadoop, 분산처리에 대한 개념까지도 꼭꼭 눌러 담은 알찬 커리큘럼. Spark를 다뤄본 경험이 없더라도 시작할 수 있습니다.

최신 업데이트 100% 반영한 최적의 실습 환경

Spark에 관심을 갖고 배워보려는 사람들은 많지만, 기초부터 체계적으로 배울 수 있는 곳은 국내 유일 패스트캠퍼스 뿐입니다. 최신 버전으로 처음부터 배우는 Spark 학습! 주목 받는 기술일수록 제대로 시작하셔야 합니다.

풍부한 경험을 갖춘 실무 전문가 강사님

현업 최전선에서 직접 빅데이터를 다루는 전문가가 여러분과 수업을 진행합니다. 실제 현업에서 사용되는 핵심 기술과 업계의 최신 트렌드, 몸으로 직접 부딪히지 않았다면 알기 어려운 실전 노하우까지! 절대 놓치지 마세요.

직접 결과물을 만들어 체득하는 실전능력

아무리 많은 지식을 쌓아도, 힘들고 지루한 코딩 연습을 수십 시간 했어도, 배운 내용을 실제 업무에 적용할 수 없다면 무슨 의미가 있을까요? 이 코스에서는 강사님과 함께 음악 추천 시스템을 기획하고 만들어 보는 시간을 갖습니다. 배운 내용, 업무에 바로 응용할 수 있는 역량을 갖추세요.



커리큘럼

- 1~3주차 ● **Apache Spark, 꼭 알아야 할 핵심 개념**
 - Apache Spark 이해하기
 - Spark 개요(Apache Spark 소개, 특징, 지원언어, Hadoop과 비교, Spark 프로그래밍 모델(RDD), 작동원리, Spark Components 등)
 - Scala 개요(Scala 언어 소개, 특징, 변수 선언, 함수 정의 등 기본 문법)
 - Spark 설치 및 실행(VirtualBox, CentOS7, Java 8, Spark 설치 및 환경설정, 프로세스, Lineage 확인 등 Spark Shell 기본 사용법)
 - Spark Core
 - Spark RDD 개요(Spark Core Components, Spark Application 배포, RDD Operations 등)
 - RDD 프로그래밍(RDD 생성, RDD Transform, RDD Action 등)
 - Shared Variables(Broadcast, Accumulators)

- 4~10주차 ● **Apache Spark의 다양한 라이브러리 정복하기**
 - Spark SQL
 - Spark SQL 개요(특징, 성능, RDD와 DataFrame/Dataset의 비교, SparkSession)
 - DataFrame/Dataset(생성법, Basic Operations, Queries, Join, 저장, Row Object 처리 등)
 - Catalyst Optimizer 및 Tungsten Project 작동원리
 - Spark Streaming
 - Spark Streaming(특징 및 기능, DStream의 개념 및 Basic Operations, Stateful Operations, Execution Model, DStream Persistence, Checkpoint, Kafka 연계 등)
 - Structured Streaming(DStream의 문제점, Programming Model, Window Operations, Watermarking 등)
 - 실습 데이터 : twitter를 이용하여 SNS 데이터 수집
 - Spark ML
 - Spark MLlib(특징, 머신러닝 분류, 데이터 유형, Word2Vec, VectorAssembler, ChiSqSelector 등)
 - MLlib Algorithms(Classification, Regression, Clustering, Collaborative, Filtering, Dimensionality Reduction)
 - ML Pipeline(Estimator, Transformer, Parameter, Pipeline)
 - Model Selection(CrossValidator, TrainValidationSplit)
 - ML Persistence(Status, Predictive Model Markup Language)

- 11~12 주차 ● **추천시스템 구현 프로젝트**
 - Recommender system with Spark
 - 추천시스템의 개요(추천시스템 아키텍처 및 알고리즘, 배치 추천시스템과 실시간 추천시스템)
 - 추천시스템 구현(음악추천시스템)
 - 실습 데이터 : 음악 예제 데이터



강사소개



강경인

- 데이터 엔지니어
- 대용량 데이터 실시간 처리/분석 기술을 연구하고 플랫폼을 개발합니다.
- 현재 통신/제조 분야에서 Apache Spark를 중심으로, 빅데이터 실시간 분석 및 머신러닝 시스템을 위한 OSS 기술지원을 담당하고 있습니다.



수강환경

강남강의장



❖ 강의에 따라 강의장이 변경될 수 있습니다.