

교육 과정 소개서.

올인원 패키지 : 실시간 빅데이터 처리를 위한
Spark & Flink Online.

안내.

해당 교육 과정 소개서는 모든 강의 영상이 촬영하기 전 작성되었습니다.

* 커리큘럼은 촬영 및 편집을 거치며 일부 변경될 수 있으나, 전반적인 강의 내용에는 변동이 없습니다.

아래 각 오픈 일정에 따라 공개됩니다.

- 1차 : 2021년 11월 15일(월)
- 2차 : 2021년 12월 08일(수)
- 2차 : 2021년 12월 29일(수)

최근 수정일자 2021년 11월 25일



강의정보

강의장	온라인 강의 데스크탑, 노트북, 모바일 등
수강 기간	평생 소장
상세페이지	https://fastcampus.co.kr/data_online_flink
담당	패스트캠퍼스 고객경험혁신팀
강의시간	20시간 예정 (* 사전 판매 중인 강의는 시간이 변경될 수 있습니다.)
문의	고객지원 : 02-501-9396 강의 관련 문의: help.online@fastcampus.co.kr 수료증 및 행정 문의: help@fastcampus.co.kr

강의특징

나만의 속도로	낮이나 새벽이나 내가 원하는 시간대에 나의 스케줄대로 수강
원하는 곳 어디서나	시간을 쪼개 먼 거리를 오가며 오프라인 강의장을 찾을 필요 없이 어디서든 수강
무제한 복습	무엇이든 반복적으로 학습해야 내것이 되기에 이해가 안가는 구간 몇번이고 재생



강의목표

- 실무 데이터 처리의 End to End 커리큘럼
데이터 처리를 위한 파이프라인을 구축하고 실무에서 그대로 사용할 수 있다.
- 우버 모빌리티 데이터로 실시간 빅데이터 처리 실습
현실 데이터로 실시간 빅데이터 처리를 직접 해보고 이해할 수 있다.
- 일괄 처리 VS 실시간 빅데이터 처리 비교 학습
각 장단점을 알고 상황에 맞게 활용할 수 있습니다.
- 오케스트레이션으로 데이터 엔지니어링을 간편하게
Apache Airflow를 활용하면 간편하게 스케줄링 가능해요

강의요약

- Apache Spark - Airflow - Flink & Kafka 3step으로 이루어져 따라오기만 하면 실무 적용 끝
Apache Spark로 데이터 병렬-분산 처리
Apache Airflow로 데이터 오케스트레이션
Apache Flink & Kafka로 데이터 스트리밍
- 우버 모빌리티 데이터로 실시간 빅데이터 처리 실습
- 열심히 구글링해도 안나오는 실무 데이터 엔지니어링백문백답



강사

김건우

과목

- 올인원 패키지 : 실시간 빅데이터 처리를 위한 Spark & Flink Online.

약력

- 현) GitHub 2.3k 팔로워 보유
(대표 레포짓 19.9k 스타)
- 현) 스타트업 Stealth 공동창업
- 전) UBER 데이터 엔지니어
- 우버 마켓플레이스 프라이싱 : 인센티브 최적화 알고리즘 백엔드 & 데이터 엔지니어링
- 우버 잇츠 프라이싱 : 배달 드라이버 공급 예측 & 배달료 프라이싱
- 우버 페이먼츠 : 비정상거래 탐지
- 우버 마켓플레이스 인센티브 : 캘리포니아 AB5법 대응 인센티브 개발 리드
- 전) 뉴욕대학교(NYC) 컴퓨터과학 전공

CURRICULUM

01.

소개

인트로
01. 데이터 엔지니어링에 대하여
02. Apache Spark
03. Apache Airflow
04. Apache Flink

본 과정은 현재 촬영 및 편집이 진행되고 있는 **사전 판매 중인 강의**입니다.
해당 교육과정 소개서는 변경되거나 추가될 수 있습니다.

CURRICULUM

02.

Apache Spark와 데이터 병렬-분산 처리

Apache Spark 시작하기

01. 빅데이터란?
02. 데이터 병렬처리와 스케일
03. 왜 Spark를 배워야 할까?
04. 데이터 저장 솔루션들
05. Apache Spark 기초
06. Spark 3

병렬처리에서에서 분산처리까지

01. RDD - Data-Parallel to Distributed Data-Parallel
02. RDD - Latency
03. RDD - RDD, 스파크의 병렬 콜렉션
04. RDD - RDD Transformation and Actions
05. 분산처리를 위한 기능들 - Reduction Operations
06. 분산처리를 위한 기능들 - Pair RDDs
07. 분산처리를 위한 기능들 - Transformations and Actions on Pair RDDs
08. 분산처리를 위한 기능들 - Joins
09. 성능 최적화 및 튜닝 방법 - Partitioning
10. 성능 최적화 및 튜닝 방법 - Shuffling
11. 성능 최적화 및 튜닝 방법 - Optimizing with Partitioners
12. 성능 최적화 및 튜닝 방법 - Wide and Narrow Dependencies

본 과정은 현재 촬영 및 편집이 진행되고 있는 **사전 판매 중인 강의**입니다.
해당 교육과정 소개서는 변경되거나 추가될 수 있습니다.

CURRICULUM

02.

Apache Spark와 데이터 병렬-분산 처리

Spark SQL 과 친구들 (DataFrames, Datasets)

01. Structured vs Unstructured Data
02. Spark SQL
03. DataFrames
04. Datasets
05. Catalyst Optimizer 및 Tungsten Project 작동원리
06. ETL with DataFrames

Spark ML 과 머신러닝 엔지니어링

01. Spark ML 기초
02. Classification and Regression using Apache Spark
03. SparkML Clustering
04. MLlib 알고리즘
05. 머신러닝 파이프라인 구성
06. 모델 평가
07. 모델 모니터링과 관리

본 과정은 현재 촬영 및 편집이 진행되고 있는 **사전 판매 중인 강의**입니다.
해당 교육과정 소개서는 변경되거나 추가될 수 있습니다.

CURRICULUM

03.

Apache Airflow와 데이터 오케스트레이션

Airflow 소개

- 01. 왜 Apache Airflow 인가?
- 02. Airflow 동작 원리
- 03. Airflow 설치
- 04. Airflow CLI
- 05. Airflow UI

나의 첫번째 데이터 파이프라인 구축하기

- 01. DAG란?
- 02. DAG 만들어보기
- 03. DAG Scheduling
- 04. Operator 알아보기 - Operators
- 05. Operator 알아보기 - Sensors
- 06. Operator 알아보기 - Transfer
- 07. Table
- 08. Backfilling and Catchup
- 09. Airflow로 Spark Pipeline 관리하기

본 과정은 현재 촬영 및 편집이 진행되고 있는 **사전 판매 중인 강의**입니다.
해당 교육과정 소개서는 변경되거나 추가될 수 있습니다.

CURRICULUM

04.

Apache Flink와 데이터 스트리밍

파트별 수강시간 00:00:00

Flink 소개
01. Flink는 무엇이 다를까
02. Flink 아키텍처와 생태계
03. Flink 설치
04. Flink 프로젝트 시작하기
데이터셋 API
01. Flink 코드 관리
02. Flink 로 단어 세기 - Map
03. Flink 로 단어 세기 - Flatmap
04. Flink 로 단어 세기 - Filter
05. Flink 로 단어 세기 - groupby
06. Joins - inner join
07. Joins - left, right & Full outer join
데이터스트림 API
01. 데이터는 어디에서 와서 어디로 갈까
02. 첫번째 데이터 스트림 만들기
03. 데이터스트림에서 데이터 처리하기 - Reduce
04. 데이터스트림에서 데이터 처리하기 - Fold
05. 데이터스트림에서 데이터 처리하기 - Aggregation
06. 데이터스트림에서 데이터 처리하기 - Split
07. 데이터스트림에서 데이터 처리하기 - Iterate
08. 스트리밍 데이터 처리하기 - Window
09. 문제 풀이하기
10. 정답 확인하고 틀린 부분 수정하기
11. 실시간 데이터 처리가 늦어질때 - Watermarks
12. 실시간 데이터 처리가 늦어질때 - Late elements
13. 시스템의 실패를 대비하여 - 장애 허용 시스템 : State, Checkpointing
[종합실습]실시간 데이터 처리해보기
1. Kafka에서 오는 데이터를 실시간 처리해보자!

본 과정은 현재 촬영 및 편집이 진행되고 있는 **사전 판매 중인 강의**입니다.
해당 교육과정 소개서는 변경되거나 추가될 수 있습니다.

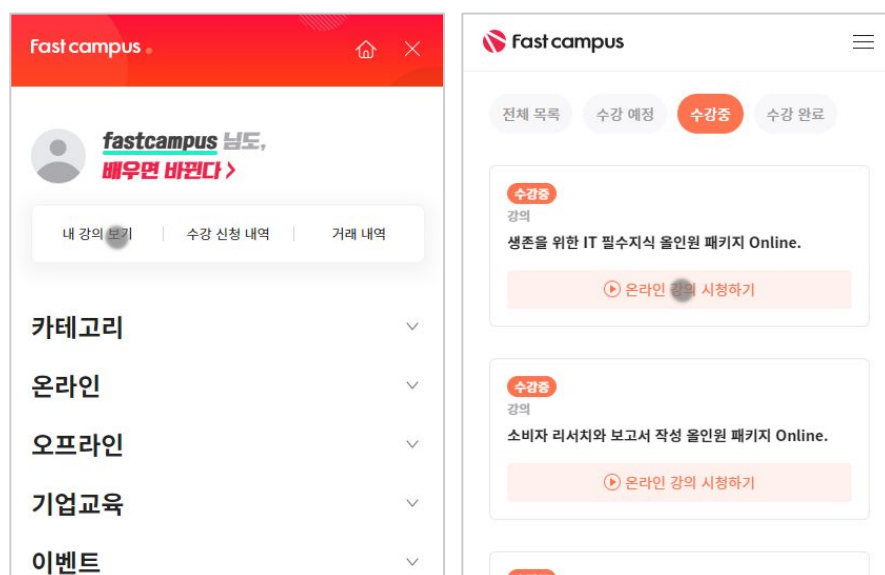


주의 사항

- 상황에 따라 사전 공지 없이 할인이 조기 마감되거나 연장될 수 있습니다.
- 패스트캠퍼스의 모든 온라인 강의는 **아이디 공유를 금지**하고 있으며 1개의 아이디로 여러 명이 수강하실 수 없습니다.
- 별도의 주의사항은 각 강의 상세페이지에서 확인하실 수 있습니다.

수강 방법

- 패스트캠퍼스는 크롬 브라우저에 최적화 되어있습니다.
- 사전 예약 판매 중인 강의의 경우 1차 공개일정에 맞춰 '온라인 강의 시청하기'가 활성화됩니다.



환불 규정

- 온라인 강의는 각 과정 별 '정상 수강기간(유료수강기간)'과 정상 수강기간 이후의 '복습 수강기간(무료수강기간)'으로 구성됩니다.
- 환불금액은 실제 결제금액을 기준으로 계산됩니다.

수강 시작 후 7일 이내	100% 환불 가능 (단, 수강하셨다면 수강 분량만큼 차감)
수강 시작 후 7일 경과	정상(유료) 수강기간 대비 잔여일에 대해 환불규정에 따라 환불 가능

※ 강의별 환불규정이 상이할 수 있으므로 각 강의 상세페이지를 확인해 주세요.