

교육 과정 소개서.

실시간 데이터 처리 위한 Spark & Flink & Kafka



강의정보

강의장	온라인 강의 데스크탑, 노트북, 모바일 등
수강 기간	평생 소장
상세페이지	https://fastcampus.co.kr/data_online_spark
강의시간	34시간 59분
문의	고객센터

강의특징

나만의 속도로	낮이나 새벽이나 내가 원하는 시간대 에 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

- Spark & Flink & Kafka & Airflow 4가지 실시간 처리 주요 프레임워크의 핵심 개념과 기술들을 배워보고 이해할 수 있습니다.
- 데이터를 생산하는 프로듀서, 데이터를 소비하는 컨슈머, 분석 및 저장하는 단계까지 실시간 데이터 처리를 위한 스케줄링과 모니터링을 통한 조직화 방법을 배울 수 있습니다.
- 현업에서 실시간 데이터 처리가 필요할 때 기술 선택부터 해결방법, 실행까지 활용하는 방법을 배울 수 있습니다.

강의요약

- Apache Spark & Flink & Kafka & Airflow 5가지 실시간 처리 주요 프레임워크의 핵심 개념과 기술 내용을 모두 담았습니다.
- 실시간 데이터를 어떻게 다루고 처리할 수 있는지 각 기술별로 다양한 예제와 실습을 진행해봅니다.
- 여러 서비스 환경과 상황에 따른 실무 데이터를 활용하여 실시간으로 데이터를 처리하는 프로젝트를 진행해봅니다.



강사

엄현호

과목

- 실시간 데이터 처리 위한 Spark & Flink & Kafka

약력

- 현) 쿠팡 data engineer

Beck

과목

- 실시간 데이터 처리 위한 Spark & Flink & Kafka

약력

- 현) 국내 대형 증권사 Data engineer
- 전) 국내 주요 포털사 Data Engineer



CURRICULUM

01.

데이터 엔지니어링이란?

파트별 수강시간 00:28:30

CH01.첫 인사 및 강의 개요
ch01-01.강의 소개 및 목차 안내
CH02.데이터 엔지니어링 소개
ch02-01.데이터 엔지니어링이란?
ch02-02.데이터 처리 관련 주요 아키텍처
CH03.실시간 데이터 처리 소개
ch03-01.실시간 데이터 처리(스트리밍)란?



CURRICULUM

02.

**배치/스트림
프로세싱 -
아파치 Spark**

파트별 수강시간 07:15:49

CH01. Spark 살펴보기
ch01-01. Apache Spark란?
ch01-02. 로컬 환경에 Spark 설치 및 워드 카운트 예제 실행
ch01-03. Spark 애플리케이션의 구성 요소
ch01-04. Transformation, Action, Lazy Evaluation의 개념
CH02. 배치 프로세싱 (Spark SQL / RDD / Dataframe)
ch02-01. Spark RDD란
ch02-02. RDD 실습 - 로그 집계 파이프라인 만들기 - map, filter, reduce, group by
ch02-03. RDD 실습 - spark UI 확인 및 Join
ch02-04. Spark DataFrame, Dataset, SQL 실습 - 로그 집계 파이프라인 만들기 - Dataframe API
ch02-05. Spark DataFrame, Dataset, SQL 실습 - 로그 집계 파이프라인 만들기 - SQL API
ch02-07. Spark DataFrame, Dataset, SQL 실습 - join
CH03. 스트림 프로세싱 (Structured Streaming, DStream)
ch03-01. Spark Structured Streaming이란
ch03-02. Structured Streaming 실습 - 실시간 로그 집계 파이프라인 만들기 Part1)
ch03-03. Structured Streaming 실습 - Join
ch03-04. DStream 실습 - read, write
ch03-05. DStream 실습 - Part 2) transformations
ch03-06. Event Time windows, Processing Time window 이론
ch03-07. Event Time windows, Processing Time window 실습
ch03-08. Watermarking 개념 및 실습
CH04. Spark 심화 개념
ch04-01. 스파크 클러스터, 런타임 아키텍처에 대한 이해
ch04-02. Deploy mode - cluster, client mode
ch04-03. Spark - action, stage, shuffle, task, slot 확인 실습
ch04-04. Join의 종류
ch04-05. 스파크에서의 메모리 할당
ch04-06. 스파크 메모리 관리(수정본)
ch04-07. Partitioning 개요 및 중요성(수정본)
ch04-08. Repartition, Coalesce에 대한 이해
ch04-09. Caching, Persistence에 대한 이해



CURRICULUM

03.

**배치/스트림
프로세싱 -
아파치 Kafka**

파트별 수강시간 02:22:38

CH01.kafka 살펴보기
ch01-01.Apache Kafka란?
CH02.kafka 기본 개념
ch02-01.Topic, Partitions, Offset
ch02-02.Producer, Message Keys
ch02-03.Consumer, Deserialization
ch02-04.Consumer Group, Consumer Offsets
ch02-05.Brokers
ch02-06.Replication
ch02-07.Zookeeper, Kraft
CH03.kafka 사용해보기
ch03-01.로컬 환경에 kafka 설치 topics cli 실습
ch03-02.producer cli 실습
ch03-03.consumer cli 실습
ch03-04.consumer group cli 실습 1)
ch03-05.consumer group cli 실습 2)
ch03-06.Producer - Java 실습
ch03-07.Producer callback - Java 실습
ch03-08.Producer with key - Java 실습
ch03-09.Consumer - Java 실습
ch03-10.Consumer - gracefully shutdown 실습

CURRICULUM

04.

스트림 프로세싱 - Apache flink

파트별 수강시간 17:03:21

CH01.Abstraction
ch01-01.다른 스트리밍 툴과 비교
ch01-02.고수준 / 저수준 API
ch01-03.Stateful 스트리밍 처리
ch01-04.Timely 스트리밍 처리
ch01-05.Architecture
ch01-06.Cluster 구현
ch01-06.Cluster 구현_EMR
ch01-06.Cluster 구현_EKS
ch01-07.HA & Execution Mode
CH02.DataStream API
ch02-01.IO (Source & Sink, Async)
ch02-02.Data Source
ch02-03.파일 Source & Sink
ch02-04.Socket Source & Sink
ch02-05.Kafka Source & Sink



CURRICULUM

04.

**스트림 프로세싱
- Apache flink**

파트별 수강시간 17:03:21

CH03.Operators
ch03-01.Data Types & Serialization
ch03-02.ProcessFunction
ch03-03.Map Operator
ch03-04.Filter Operator
ch03-05.FlatMap Operator
ch03-06.KeyBy Operator
ch03-07.Reduce Operator
ch03-08.Windows
ch03-09.ProcessWindowFunction
ch03-10.Window Operator
ch03-11.Tumbling Window
ch03-12.Sliding Window
ch03-13.Session Window
ch03-14.Global Window
ch03-15.Joining
ch03-16.Tumbling Window Join
ch03-17.Sliding Window Join
ch03-18.Session Window Join
ch03-19.Interval Join
ch03-20.Connect Operator
ch03-21.CoMap
ch03-22.CoFlatMap



CURRICULUM

04.

**스트림 프로세싱
- Apache flink**

파트별 수강시간 17:03:21

CH04.State & Fault Tolerance
ch04-01.Keyed State
ch04-02.ValueState
ch04-03.ReducingState
ch04-04.Iterate
ch04-05.AggregationState
ch04-06.MapState
ch04-07.Broadcast State 패턴
ch04-08.Checkpointing - Savepoints
ch04-09.Queryable State
ch04-10.데이터 타입 & 직렬화
ch04-11.Side Outputs
CH05.Event Processing(CEP)
ch05-01.Pattern API (Individual)
ch05-01.Pattern API (Combining)
ch05-02.단일 / 결합 패턴
ch05-03.패턴 그룹 & Skip 전략
ch05-04.Time in CEP
CH06.Flink ML
ch06-01.Overview
ch06-02.Pipeline & Graph
ch06-03.Vector
ch06-04.Clustering
CH07.Deployment
ch07-01.Memory Configuration(Component)
ch07-01.Memory Configuration(Tuning)
ch07-02.Elastic Scaling
ch07-03.Fine-Graded Resource Management
ch07-04.Metric Reporters / Logging
CH08.Semi Project
ch08-01.Semi Project



CURRICULUM

05.

Airflow 통한 배치 프로세싱

파트별 수강시간 06:40:07

CH01.Abstraction
ch01-01.Architecture
ch01-02.Install & UI
ch01-03.CLI
CH02.DAG
ch02-01.DAG
ch02-02.Tasks
ch02-03.Control Flow
ch02-04.Timetables
ch02-05.Catchup & Backfill
ch02-06.Timeout
ch02-07.Callbacks & Notifier
ch02-08.Pools
ch02-09.Connections & Hook
ch02-10.SubDAGs & Hook
ch02-11.XComs & Variables
ch02-12.Params
ch02-13.Configuration
CH03.Operator
ch03-01.PythonOperator
ch03-02.BashOperator
ch03-03.PostgresOperator
ch03-04.BranchOperator
ch03-05.BranchDateTimeOperator
ch03-06.SubDagOperator & TaskGroup
ch03-07.TriggerDagRunOperator
ch03-08.ExternalTaskSensor
ch03-09.LastestOnlyOperator
ch03-10.Misc Operator & Seneor
ch03-11.BaseOperator
ch03-12.CustomOperator



CURRICULUM

06.

**스트림 프로세싱
프로젝트**

파트별 수강시간 01:09:09

CH01.프로젝트 개요
ch06-01.프로젝트 개요
CH02.데이터 프로듀싱
ch02-01.Producer - naver,kakao도서 api 작성
ch02-02.protocol buffer, topic, Producer
CH03.데이터 스트리밍
ch03-01.Consumer - Spark Streaming

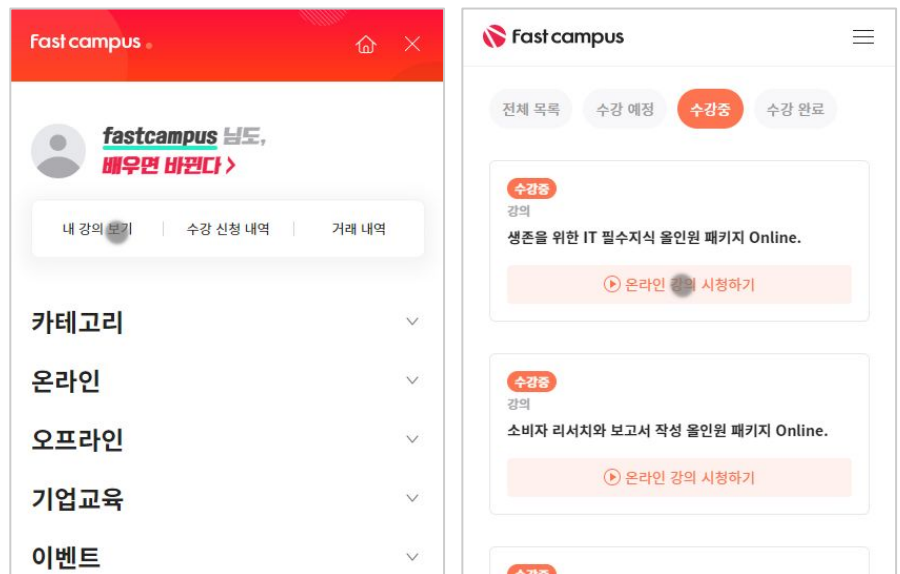


주의 사항

- 상황에 따라 사전 공지 없이 할인이 조기 마감되거나 연장될 수 있습니다.
- 패스트캠퍼스의 모든 온라인 강의는 아이디 공유를 금지하고 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.