

# 교육 과정 소개서.

---

실시간 빅데이터 처리를 위한 Spark & Flink Online.



## 강의정보

강의장	온라인 강의   데스크탑, 노트북, 모바일 등
수강 기간	평생 소장
상세페이지	<a href="https://fastcampus.co.kr/data_online_flink">https://fastcampus.co.kr/data_online_flink</a>
담당	패스트캠퍼스 고객경험혁신팀
강의시간	18시간 4분
문의	고객지원 : 02-501-9396 강의 관련 문의: <a href="mailto:help.online@fastcampus.co.kr">help.online@fastcampus.co.kr</a> 수료증 및 행정 문의: <a href="mailto:help@fastcampus.co.kr">help@fastcampus.co.kr</a>

## 강의특징

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



## 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.

## 들어가며

파트별 수강기간 : 00:22:07

---

CH01_01. 데이터 엔지니어링이란
CH01_02. 모던 데이터 엔지니어링 아키텍처
CH01_03. Batch & Stream Processing
CH01_04. Dataflow Orchestration
CH01_05 데이터 엔지니어링 프로젝트 소개



CURRICULUM

02.

Apache Spark와  
데이터 병렬-분산  
처리

파트별 수강기간 : 06:51:21

<b>Chapter 01. Apache Spark 시작하기</b>
CH01_01. 기초 환경 설정 (맥)
CH01_02. 기초 환경 설정 (윈도우)
CH01_03. 모빌리티 데이터 다운로드
CH01_04. 우버 트립수 세기
CH01_05. Spark에 대해 알아보자
CH01_06. Spark는 어떻게 진화하고 있을까
CH01_07. Resilient Distributed Dataset (RDD)
CH01_08. 코드 한줄씩 파해치기
<b>Chapter 02. 병렬처리에서 분산처리까지</b>
CH02_01. 병렬처리와 분산처리
CH02_02. 분산처리와 Latency_1
CH02_03. Key-Value RDD
CH02_04. RDD Transformations and Actions
CH02_05. Cache & Persist
CH02_06. Cluster Topology
CH02_07. Reduction Operations
CH02_08. Key-Value RDD Operations & Joins
CH02_09. Shuffling & Partitioning
<b>Chapter 03. Spark SQL, Dataframe, 그리고 Dataset</b>
CH03_01. Structured vs Unstructured Data
CH03_02. Spark SQL 소개
CH03_03. SQL 기초
CH03_04. DataFrame
CH03_05. Spark SQL 로 트립 수 세기
CH03_06. Spark SQL로 뉴욕의 각 행정구 별 데이터 추출하기
CH03_07. Catalyst Optimizer 및 Tungsten Project 작동원리
CH03_08. User Defined Functions
CH03_09. 뉴욕 택시 데이터 분석




---

CURRICULUM

02.

**Apache Spark와  
데이터 병렬-분산  
처리**

파트별 수강기간 : 06:51:21

---

<b>Chapter 04. Spark ML과 머신러닝 엔지니어링</b>
CH04_01. MLlib과 머신러닝 파이프라인 구성
CH04_02. 첫 파이프라인 구축
CH04_03. 추천 알고리즘
CH04_04. 영화 추천 파이프라인 구축
CH04_05. Regression & Classification
CH04_06. 택시비 예측하기 1
CH04_07. 택시비 예측하기 2
CH04_08. 하이퍼파라미터 최적화
CH04_09. 모델 저장 & 로딩
<b>Chapter 05. Spark Streaming</b>
CH05_01. Spark Streaming
CH05_02. 간단한 스트리밍 구현



CURRICULUM

03.

**Apache  
Airflow와 데이터  
오케스트레이션**

파트별 수강기간 : 02:27:45

<b>Chapter 01. Apache Airflow 알아보기</b>
CH01_01. 워크플로우 관리를 위한 에어플로우
CH01_02. Airflow의 구조
CH01_03. Airflow 설치
CH01_04. Airflow CLI
CH01_05. Airflow UI
<b>Chapter 02. 나의 첫번째 데이터 파이프라인 구축하기</b>
CH02_01. NFT 파이프라인 프로젝트 소개
CH02_02. DAG Skeleton
CH02_03. Operators
CH02_04. Table 만들기
CH02_05. Sensor로 API 확인하기
CH02_06. HttpOperator로 데이터 불러오기
CH02_07. PythonOperator로 데이터 처리하기
CH02_08. BashOperator로 데이터 저장하기
CH02_09. 태스크간 의존성 만들기
CH02_10. Backfill
<b>Chapter 03. Airflow로 Spark 파이프라인 관리하기</b>
CH03_01. Airflow로 Spark 파이프라인 관리하기 - Airflow와 Spark 같이쓰기
CH03_02. 택시비 예측 파이프라인 만들기 1
CH03_03. 택시비 예측 파이프라인 만들기 2





CURRICULUM

04.

**Apache Flink & Kafka와 실시간 빅데이터 처리**

파트별 수강기간 : 02:43:13

<b>Chapter 01. Kafka 소개</b>
CH01_01. Kafka에 대해 알아보자
CH01_02. Kafka의 구조
CH01_03. Topics, Partitions, 그리고 Messages
CH01_04. Cluster & Replication
CH01_05. Producer
CH01_06. Consumer
CH01_07. Zookeeper
CH01_08. 카프카 설치
<b>Chapter 02. Kafka CLI</b>
CH02_01. Zookeeper 콘솔에서 사용하기
CH02_02. Broker 시작하기
CH02_03. Topic 만들기
CH02_04. Producer CLI
CH02_05. Consumer CLI
CH02_06. Consumer Group CLI
CH02_07. Consumer와 Partitions
<b>Chapter 03. Kafka 프로그래밍</b>
CH03_01. 파이썬과 Kafka 기초
CH03_02. 파이썬으로 Producer 만들기
CH03_03. 파이썬으로 Consumer 만들기
CH03_04. 도커 설치
CH03_05. Kafka 클러스터 구축 준비
CH03_06. Kafka 클러스터 구축하기-Zookeeper
CH03_07. Kafka 클러스터 구축하기-Kafka Broker
CH03_08. Kafka 클러스터 구축하기-Kafdrop
CH03_09. Kafka 클러스터 구축하기-Topic 만들기
CH03_10. Kafka 클러스터를 이용하는 Producer
CH03_11. Kafka 클러스터를 이용하는 Consumer
CH03_12. Kafdrop으로 메시지 확인하기
CH03_13. CSV를 스트림으로 바꿔주는 Producer
CH03_14. 비정상 데이터 탐지-Producer
CH03_15. 비정상 데이터 탐지-Detector
CH03_16. 비정상 데이터 탐지-Processor



CURRICULUM

05.

**Apache Flink와  
스트리밍  
프로세싱**

파트별 수강기간 : 05:40:06

<b>Chapter 01. Apache Flink 알아보기</b>
CH01_01. Flink 소개
CH01_02. Stream Processing
CH01_03. Hadoop vs Spark vs Flink
CH01_04. Flink 구성과 생태계
CH01_05. Flink 프로그래밍
CH01_06. Stateful Stream Processing
CH01_07. Timely Stream Processing
CH01_08. Flink 아키텍처
CH01_09. pyflink
CH01_10. Flink 설치
CH01_11. 단어 수 세기
<b>Chapter 02. Source와 Sink</b>
CH02_01. Table API
CH02_02. Table Environment 만들기
CH02_03. Source
CH02_04. Source Table 만들기
CH02_05. CSV Source Table 만들기
CH02_06. Stream환경에서 CSV Source Table 만들기
CH02_07. Kafka Source Table 만들기
CH02_08. SQL로 카프카와 연결하기
CH02_09. Sink
CH02_10. CSV Sink Table 만들기
CH02_11. Kafka Sink Datastream
CH02_12. Kafka Sink Table 만들기
<b>Chapter 03. Table API와 SQL</b>
CH03_01. Explain
CH03_02. 여러개의 Sink를 동시에 만들기
CH03_03. DataStream과 Table 같이 쓰기
CH03_04. 끝이 있는 테이블과 다이내믹 테이블
CH03_05. Table Operation
CH03_06. Select
CH03_07. Filter
CH03_08. Join
CH03_09. Group By



CURRICULUM

05.

**Apache Flink와  
스트리밍  
프로세싱**

파트별 수강기간 : 05:40:06

<b>Chapter 04. DataStream</b>
CH04_01. 스트림 데이터 모아보기 - Windows
CH04_02. Tumbling vs Sliding vs Session Windows
CH04_03. Event Time과 Watermark
CH04_04. Table API를 사용한 Tumbling Windows 구현
CH04_05. SQL을 사용한 Tumbling Windows 구현
CH04_06. Table API를 사용한 Sliding Windows 구현
CH04_07. SQL을 사용한 Sliding Windows 구현
CH04_08. Table API를 사용한 Session Windows 구현
CH04_09. SQL을 사용한 Session Windows 구현
CH04_10. State 접근
CH04_11. UDF
<b>Chapter 05. Twitter 분석</b>
CH05_01. Twitter 개발자 등록 & 개발 환경 설정
CH05_02. Twitter Stream Producer
CH05_03. Twitter Stream 단어 수 세기
<b>Chapter 06. Taxi Pricing</b>
CH06_02. Flink로 택시 정보 받기
CH06_03. 택시비 예측하기

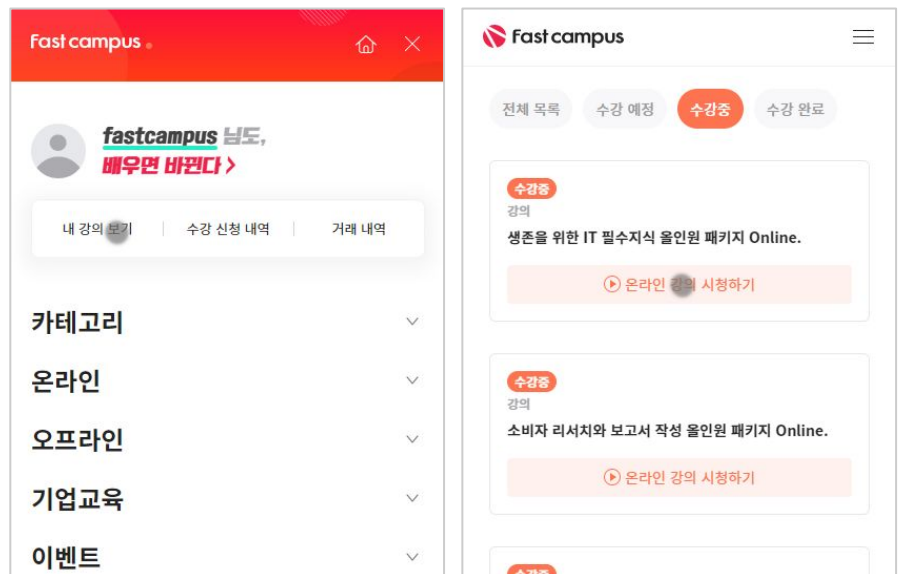


## 주의 사항

- 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.