

교육 과정 소개서.

실무에 활용하는 데이터 파이프라인 구축 CAMP



코스요약

코스명	실무에 활용하는 데이터 파이프라인 구축 CAMP
기간	2022년 7월 13일 ~ 8월 17일(총 6주)
일정	수요일 19:00 ~ 22:30(주 1회/총 21시간)
장소	비대면 원격 강의(ZOOM)
준비물	ZOOM 강의를 시청할 수 있는 온라인 환경/ 개인 랩탑
담당자	02-501-9396 / help-gathering@fastcampus.co.kr
수강료	1,000,000
상세페이지 url	fastcampus.co.kr/data_camp_pipeline

코스목표

본 CAMP에서는 현업 경력 20년의 베테랑 강사님과 함께 데이터 파이프라인 구축 실습을 진행하고, 실제 서비스 배포와 운영 노하우까지 단 6주안에 모두 배우실 수 있습니다.

코스정보

데이터 파이프라인 구축 실습으로 실전력을 갖추고, 실제 운영 노하우까지! 실제 데이터가 발생하는 환경 상에서 실습 진행. 실무에서 바로 데이터 파이프라인을 구축할 수 있도록. 본 강의에서 배우실 수 있습니다.



코스특징

빠르게 데이터 파이프라인을 구축하는 방법을 배울 수 있습니다.

본 강의는 Apache Spark와 AWS를 활용하여 직접 데이터 파이프라인을 구축하는 실습형 강의입니다. 데이터 수집부터 데이터 저장, 전처리, 분석과 시각화까지 모두 배워 실제 서비스에 바로 적용할 수 있습니다.

AWS의 경제적인 운영 방법과 서비스 제공 노하우를 배울 수 있습니다.

AWS는 유료 서비스이기 때문에 경제적으로 활용하는 것이 중요합니다. 하지만, AWS에서는 이러한 노하우를 가르쳐주지 않습니다. AWS를 많이 활용한 현업 전문가에게 직접 AWS의 활용 노하우를 배워가세요.

AWS와 Spark의 사용법이라는 두 마리 토끼를 잡을 수 있습니다.

본 강의를 통해 AWS와 Spark이 갖고 있는 다양한 기능들을 어떻게 조합해야 하는지에 대한 감을 잡을 수 있습니다. 또한 데이터웨어하우스와 데이터 레이크 전체에 걸쳐 간단하게 모든 데이터를 분석할 수 있는 Redshift를 어떻게 활용하는지까지 배울 수 있습니다.

미니프로젝트를 진행해보며, 강의 전체 내용을 wrap-up 해 볼 수 있습니다.

미니프로젝트를 통해 데이터를 분석할 Target을 선정하고, 데이터 구성 및 수집부터 데이터 가공까지 직접 경험할 수 있습니다. 또한 미니프로젝트 결과물을 발표하는 시간을 가지며 다른 수강생들과 결과를 공유해볼 수 있습니다.



커리큘럼

- 1회차 ● 데이터 수집**
 - 데이터 파이프라인의 흐름의 이해
 - 데이터 파이프라인의 각 단계에서 사용되는 최신 패키지
 - 데이터 수집을 위한 패키지인 logstash와 kafka를 이용해 ec2에서 로그를 s3에 수집하기
 - 미니프로젝트 주차별 진행 예정

- 2회차 ● AWS 패키지들을 이용한 데이터 수집**
 - Api-gateway, Kinesis Stream, Firehose, S3의 이해
 - Api-gateway, Kinesis Stream, Firehose, S3 설정하기
 - AWS cloud 에서 데이터 수집하기
 - 미니프로젝트 : 데이터 분석할 Target 설정하기

- 3회차 ● 데이터 저장 및 전처리(1)**
 - EMR 구조에 대한 이해 및 실행
 - EMR의 Spark에 대한 모니터링 방법의 이해
 - Zeppelin을 이용한 데이터 처리/ Zeppelin에서 RDS와 Glue에 정의된 테이블 join 해보기
 - 미니프로젝트 : 데이터 구성 및 수집하기

- 4회차 ● 데이터 저장 및 전처리(2)**
 - AWS Glue의 이해/ AWS Glue를 활용한 Hive table 생성
 - Hive table를 활용한 데이터 처리
 - Zeppelin의 cron 기능을 활용한 데이터 생성
 - 미니프로젝트 : 데이터 가공하기

- 5회차 ● 데이터 분석/시각화**
 - Presto 이용해 S3 저장되어 있는 Hive Table Data 를 BI툴에 제공하기
 - Presto와 Tableau를 연동하여 간단한 Dashboard 구성해보기
 - Spark와 ElasticSearch 와 연동해서 간단한 분석 수행해보기
 - 미니프로젝트 : 시각화(Visualization)하기

- 6주차 ● 데이터 저장 및 전처리(3)**
 - Redshift의 이해/ Redshift SQL 조회
 - Redshift Spectrum을 이용해서 Glue에 정의되어 있는 Table 조회
 - Snowflake의 이해/ Snowflake hands on
 - 미니프로젝트 : 미니프로젝트 결과를 발표



강사소개



권낙주

현) S은행 인공지능 플랫폼 책임자

[약력]

인덕대학교 겸임교수

전) SK C&C AccuInsight Public Cloud 아키텍트 책임자

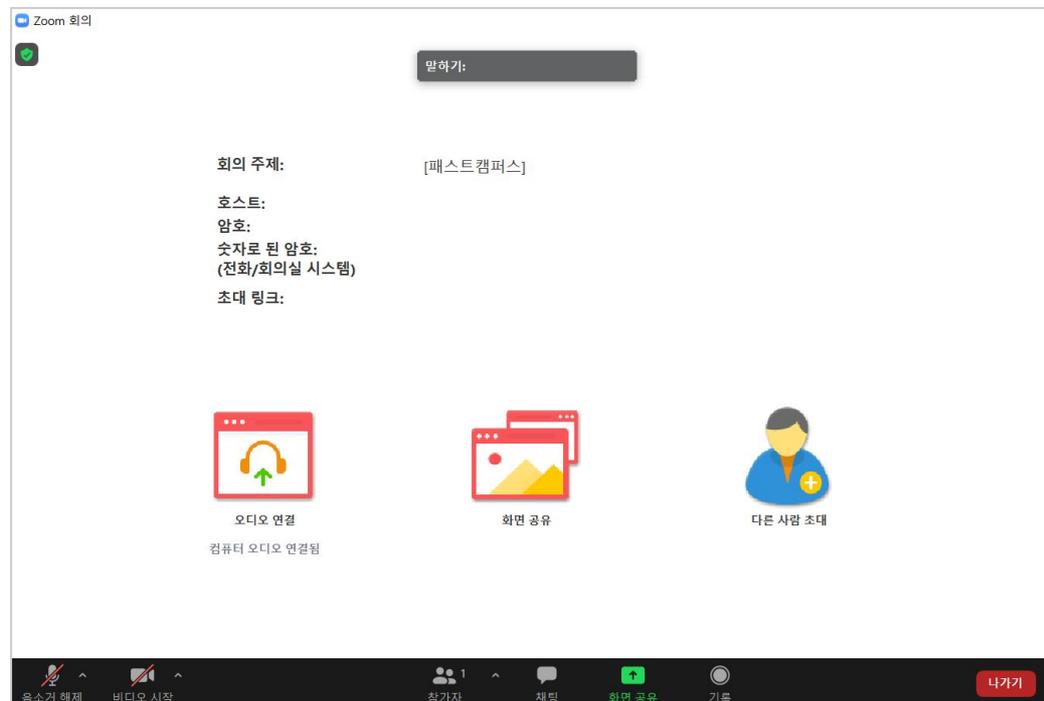
전) 직방 데이터분석팀 데이터 아키텍트 책임자 금융권 데이터표준화 컨설팅

솔루션 개발 기술자문 10년 경력



수강환경

본 강의는 '비대면 원격 강의(ZOOM)'으로 진행됩니다.



❖ 강의에 따라 강의장이 변경될 수 있습니다.