1

Cloud 인프라 구축
데이터 파이프라인 구축을 위한 기반이 되는 Cloud 환경 구성을 위한 툴을 배우고 실습합니다. Amazon Web Services(이하 AWS)는 유료 서비스이기 때문에 경제적으로 활용하는 것이 중요합니다. AWS를 많이 활용한 현업 전문가에게 직접 AWS의 활용 노하우를 배워가세요.

2

데이터 엔지니어링 개요
본격적인 데이터 엔지니어링을 위한 전체적인 개요를 꼼꼼하게 학습합니다. 빅데이터를 배우기 앞서 데이터 분야의 전반적인 개요와 함께 기존 관계형 데이터베이스(RDBMS) SQL을 맛보고, 실제 현업에서의 Case Study를 통해 실제 데이터 엔지니어들의 실무를 배웁니다.

3

Python
Python에는 목적에 맞는 데이터 분석을 손쉽게 할 수 있도록 Numpy, Pandas, Matplotlib와 같은 다양한 라이브러리가 개발되어 있습니다. 데이터 수집 뿐 아니라 가장 까다로운 전처리, 그리고 시각화까지 각 단계별로 활용할 수 있는 Python 라이브러리를 다양한 실습을 통해 깊이있게 학습할 수 있습니다. Spark 2.0에서 더욱 강력해진 Python 언어 지원으로, 엔지니어링 뿐만 아니라 분석 능력까지, 두 마리 토끼를 잡아보세요!

4

Spark
Apache Spark의 정의와 특징을 살펴보고, 작동 원리를 이해하고, 기본적인 사용법과 실습 진행을 위한 설치, 환경설정을 배웁니다. 나아가 Core Components와 DataFrame 학습을 통해 Spark SQL 뿐만 아니라 스트리밍, 머신러닝까지! 다양한 라이브러리를 정복할 수 있습니다.

5

클라우드 플랫폼을 활용한 실습
AWS와 Spark이 갖고 있는 다양한 기능들을 어떻게 조합해야 하는지에 대한 감을 잡을 수 있습니다. 또한 AWS와 Spark를 활용하여 데이터 파이프라인의 실제 운영 방법과 트러블 슈팅까지 배울 수 있습니다. 데이터 파이프라인 구축의 핵심적인 AWS 툴들(GLUE, EMR, Redshift, Kinesis, Elasticserach, Sage Maker 등) 뿐 만 아니라, Google Cloud Platform Bigquery 실습을 통한 폭넓은 클라우드 플랫폼을 다룰 수 있습니다.

6

시각화
데이터의 양상과 분석된 결과값을 직관적으로 파악하고, 인사이트를 얻기 위해서는 '시각화'가 필수입니다. 데이터 시각화의 대세주자 Tableau, 라는 간편한 시각화 툴을 활용해 별도 프로그래밍 없이 손쉽게 그래프와 대쉬보드를 제작하는 법을 배우고, 또한 Google Data Studio 실습을 통해 다양한 방법으로 정형/비정형 데이터를 시각화해 봅니다.

7

머신러닝, 딥러닝
Spark MLlib을 활용한 머신러닝 데이터 분석 실습을 통해, 머신러닝 엔지니어들과 소통할 수 있을 정도의 라이브러리 및 메소드 들의 기능을 학습합니다. 프로젝트를 통해 내 손으로 직접 머신러닝/딥러닝 모델을 구현해볼 수 있습니다.

8

프로젝트 (파이프라인 구축 / 머신러닝 / 개인프로젝트)
수업시간에 배운 실습내용을 토대로 가이드 프로젝트 형태의 데이터 파이프라인 구축과 머신러닝 모델 구현 프로젝트를 진행합니다. 파이프라인 구축, 머신러닝을 통한 분석 프로젝트 이후엔, 개인의 입맛에 맞게 데이터 셋을 준비, 분석, 결과 도출까지의 전체적인 과정을 거치며, 강사님의 피드백을 바탕으로 나만의 프로젝트가 완성됩니다!

클라우드 플랫폼 실습 강의

데이터, 분석 이론 및 Spark 강의

주차	회차 및 날짜	소주제	세부 내용
1	1회차 2/19 화	Cloud_AWS 1	- 클라우드 서비스 이해 - AWS 이해 - AWS EC2 이해 - VPC 생성 - IAM 권한관리
	2회차 2/23 토	Data Engineering Intro	- 데이터 분야 소개 - 컴퓨터 엔지니어링 분야 소개 - 데이터 엔지니어링? - 머신러닝 & 머신러닝 엔지니어링 소개 - Case Study
2	3회차 2/26 화	Cloud_AWS 2	- EC2 인스턴스 생성 - EC2 관리 실습 - 무제한 용량의 객체 스토리지 : S3 - S3 버킷 생성, 관리, 설정 - CloudWatch 를 이용한 모니터링
	4회차 3/2 토	Python for Spark	- Python
3	5회차 3/5 화	RDBMS 실습 및 효율화	- Managed RDBMS: RDS - RDS 를 이용해 SQL기반의 데이터 분석 실습 - 기존의 방식으로 성능을 올리려면? -- Scale Up -- 부하 분산
	6회차 3/9 토	Big Data Intro	- 기존의 데이터 시스템 (RDBMS) - 한계 - GFS, MapReduce Concept - Hadoop - MapReduce - Hive - 분산 데이터베이스 - 클라우드 제품들
4	7회차 3/12 화	Spark Intro	- Apache Spark 개요 - Spark core concept - RDD, DataFrame, Machine Learning - Spark Demo
	8회차 3/16 토	Spark 실습1 - DataFrame 기본	- Spark DataFrame 개요 - Spark DataFrame 데이터 분석 실습 - Spark SQL
5	9회차 3/19 화	Data Engineering 실습 1 (전처리 및 저장)	- AWS GLUE - EMR - Athena - Redshift
	10회차 3/23 토	Spark 실습2 - DataFrame 심화	- Spark DataFrame UDF - Spark DataFrame Analytic Function
6	3/26 화	BREAK (3/26 화요일)	
	11회차 3/30 토	Spark Cluster	- 클러스터 구축 리뷰 (EMR) - Master, Slave Script - Cluster 구조 (이론) - Cluster UI
7	12회차 4/2 화	Data Engineering 실습 2 (수집)	- Kinesis stream - Kinesis Firehose - Lambda
	13회차 4/6 토	Spark 실습3 - Spark Streaming 1	- Spark Streaming 이론 수업 - 실습환경 구축 - 트위터 실시간 분석 실습
8	14회차 4/9 화	Data Engineering 실습 3 (Dashboard)	- AWS ES(Elasticsearch Service) - DMS(Database Migration Service)
	15회차 4/13 토	가이드 프로젝트 1	- 가이드 프로젝트 1 진행 (1) - 실시간 데이터 파이프라인 - 트위터 스트림 - Kinesis - Structured Streaming

클라우드 플랫폼 실습 강의

데이터, 분석 이론 및 Spark 강의

주차	회차 및 날짜	소주제	세부 내용
9	16회차 4/16 화	**Cloud_GCP**	- AWS 와 GCP(Google Cloud Platform) 용어비교 - GCP Bigquery
	17회차 4/20 토	**시각화**	- 여러가지 시각화 방법 - BI Tools (Dashboard) - Tableau, DataStudio
10	18회차 4/23 화	**Sage Maker for ML**	- Sage Maker
	19회차 4/27 토	**가이드 프로젝트 1**	- 가이드 프로젝트 1 진행 (2)
11	20회차 4/30 화	**ML Intro**	- 머신러닝 개요
	21회차 5/4 토	**ML 실습**	- 머신러닝 실습 (MLlib, Scikit Learn)
12	22회차 5/7 화	**DL Intro**	- 딥러닝 개요 - 딥러닝을 위한 인프라
	23회차 5/11 토	**가이드 프로젝트 1**	- 가이드 프로젝트 1 진행 (3)
13	24회차 5/14 화	**DL 실습**	- 딥러닝 실습 (TensorFlow)
	25회차 5/18 토	**가이드 프로젝트 2 (with ML)**	- 가이드프로젝트 2 진행 (1) - 실시간 데이터 파이프라인 - 머신러닝, 딥러닝 모델 학습 - 머신러닝, 딥러닝 모델 적용
14	26회차 5/21 화	**가이드 프로젝트 2 (with ML)**	- 가이드 프로젝트 2 진행 (2)
	27회차 5/25 토	**가이드 프로젝트 2 (with ML)**	- 가이드 프로젝트 2 진행 (3)
15	28회차 5/28 화	**개인 프로젝트**	- 개인 프로젝트 진행 (1)
	29회차 6/1 토	**개인 프로젝트**	- 개인 프로젝트 진행 (2)
16	30회차 6/4 화	**개인 프로젝트**	- 개인 프로젝트 진행 (3)