

● 교과목 개요

정보화 시대에 적합한 오픈 소스 데이터 처리 기법을 학습함을 목표로 한다.

이를 위하여 오픈 소스 데이터 처리 도구(spark)를 직접 실습할 기회와 더불어 정보기술의 기본 지식이 데이터 구조와 Python 프로그래밍 언어에 대하여 소개한다.

● 수업목표

Spark를 사용하여 오픈 소스 환경에서 데이터 처리 과정을 실습해 봄으로써,

- 데이터 처리에 대한 이해 능력을 향상하고,
- 실무 처리 능력을 배양한다.

이를 위하여

- 데이터 구조에 대한 기본 개념들을 소개하고
- Spark 실행을 위한 언어로서 python을 사용한 기초적인 프로그래밍 기술을 학습한다.

● 주차별 강의 계획서

1주. 과목 소개 및 Spark thro

2주. Spark 설치 및 실행 환경 구성

3주. 파이썬 프로그래밍 소개

4주. 파이썬 자료구조 소개

5주. PySpark 변환함수 및 액션 소개 및 실행

6주. Matplotlib를 사용한 파이썬 그래프 작성

7주. DataFrame 과 SparkSQL 소개 및 실행 (단일 데이터프레임 처리)

8주. (중간 시험)

9주. DataFrame 과 SparkSQL 소개 및 실행 (데이터프레임 조인 처리)

10주. PySpark 변환함수 및 액션 추가 설명

11주. PySpark DStream 설명 및 실습

12주. PySpark StructuredStreaming 설명 및 실습

13주. BigData 아키텍처

14주. 최종 정리

15주. (기말 시험)