

# RDD 에서 Dataframe 생성 방법

## I. 비 정형 데이터인 로그 데이터

RDD로 로드하여 각 줄을 파싱 → 정형화 과정 → Dataframe으로 활용

### RDD에서 Dataframe을 생성하는 방법 3가지

- 1) 로우의 데이터를 튜플 형태로 저장한 RDD를 사용하는 방법
- 2) 케이스 클래스를 사용하는 방법
- 3) 스키마를 명시적으로 지정하는 방법

## II. 정형 데이터 Source Dataframe으로 불러오기

SQL DB, JDBC, ODBC, Mongo DB, My SQL DB 등 커넥터 라이브러리 제공

✓ (기존 RDD에서 생성한 Spark Context가 아닌) Dataframe 및 SQL API는 **SparkSession** 객체

'sparkSession'으로  
Dataframe과 RDD를 다  
생성 가능하다.