

# IPython과 Pandas를 활용한 게임 데이터 분석



# 발표자 소개

- 김정주(@haje01)
- 원 3D엔진+게임 클라이언트 개발
- 현재 'Webzen & Play'에서 게임 데이터 수집 및 분석 시스템 개발

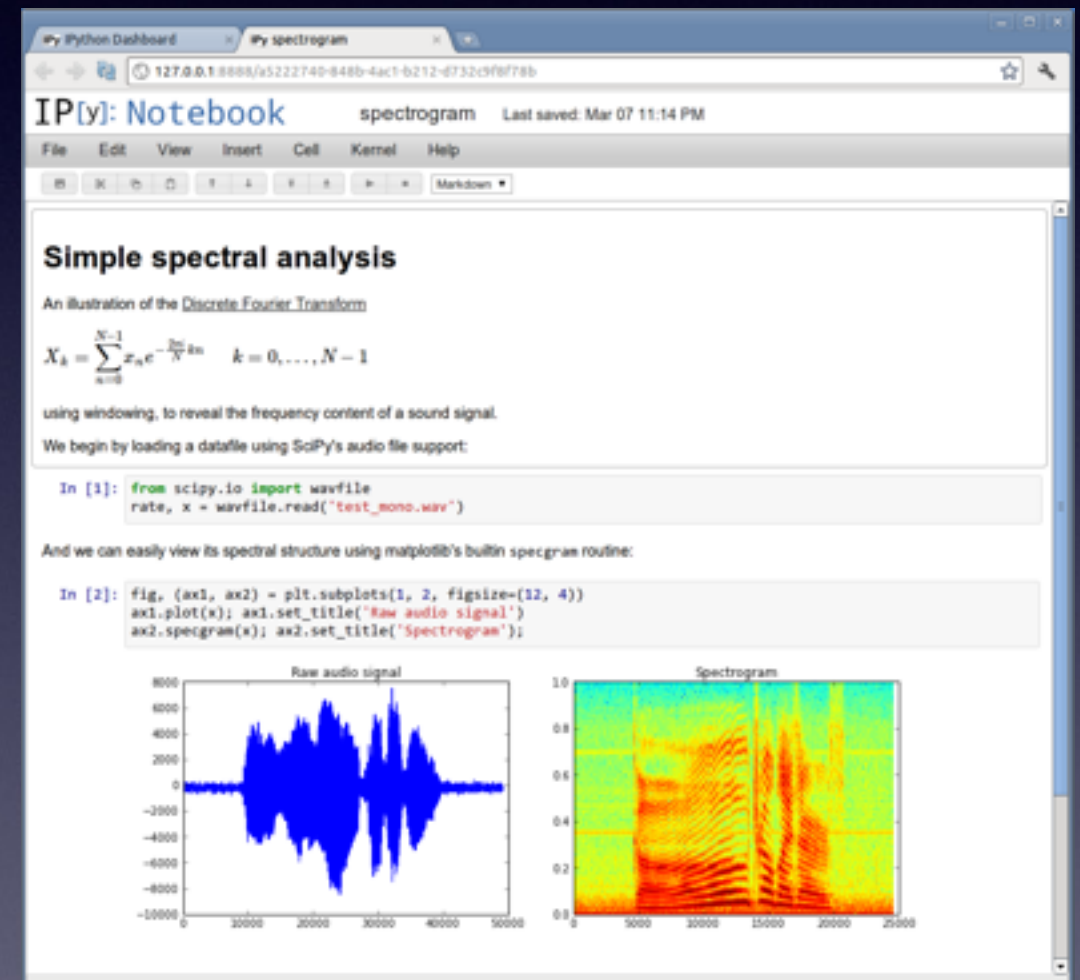


# IPython과 Pandas



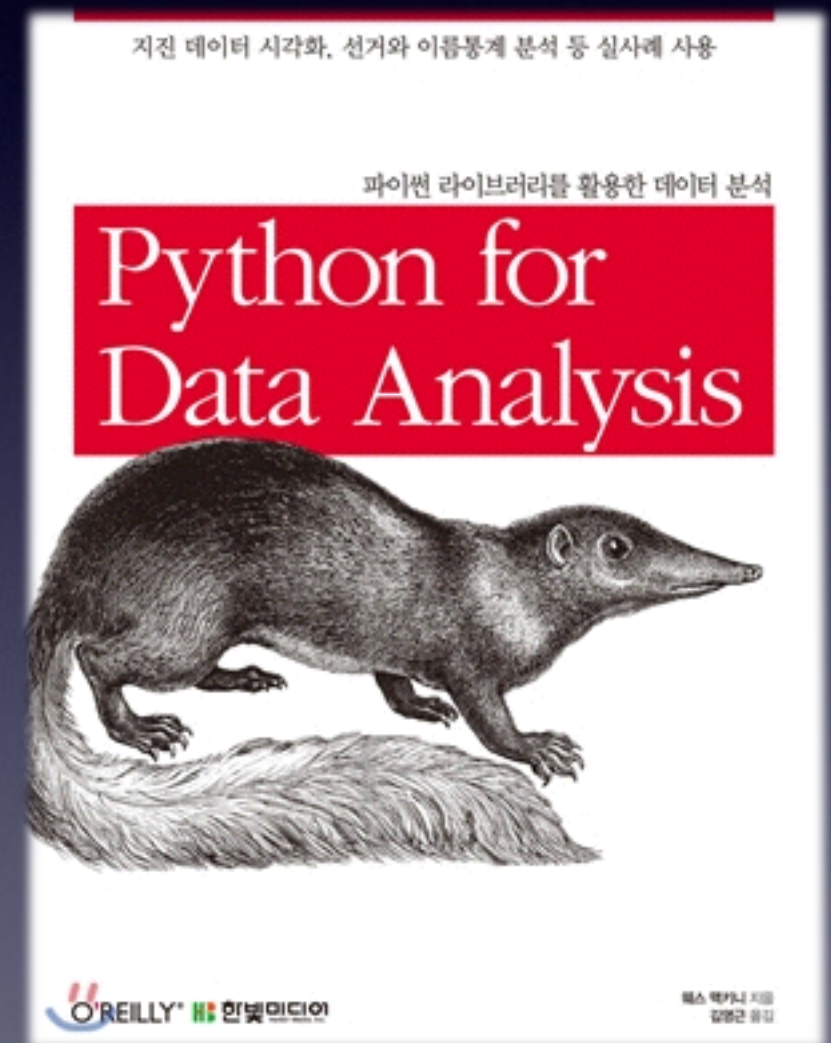
# IPython (notebook)

- 파이썬을 위한 더 나은 인터랙티브 컴퓨팅 환경
- 웹브라우저 기반으로 Try & Discover를 편리하게 할 수 있다. (데이터 분석에 최적)



# Pandas

- R의 분석 능력을 파이썬으로
- 데이터의 정형화(Munging)
  - 주로 .csv 파일을 가져와 가공
- 강력한 시계열 분석(TimeSeries)
- 통계 및 분석 관련 다양한 기능들



WzDat



# WzDat?

- Webzen Data Toolkit의 약자(‘왓츠뎃’으로 읽음)
- IPython + Pandas 기반의 게임 (로그)데이터 분석 시스템을 만들기 위해 시작
- IPython과 Pandas만으로 부족한 기능을 보강

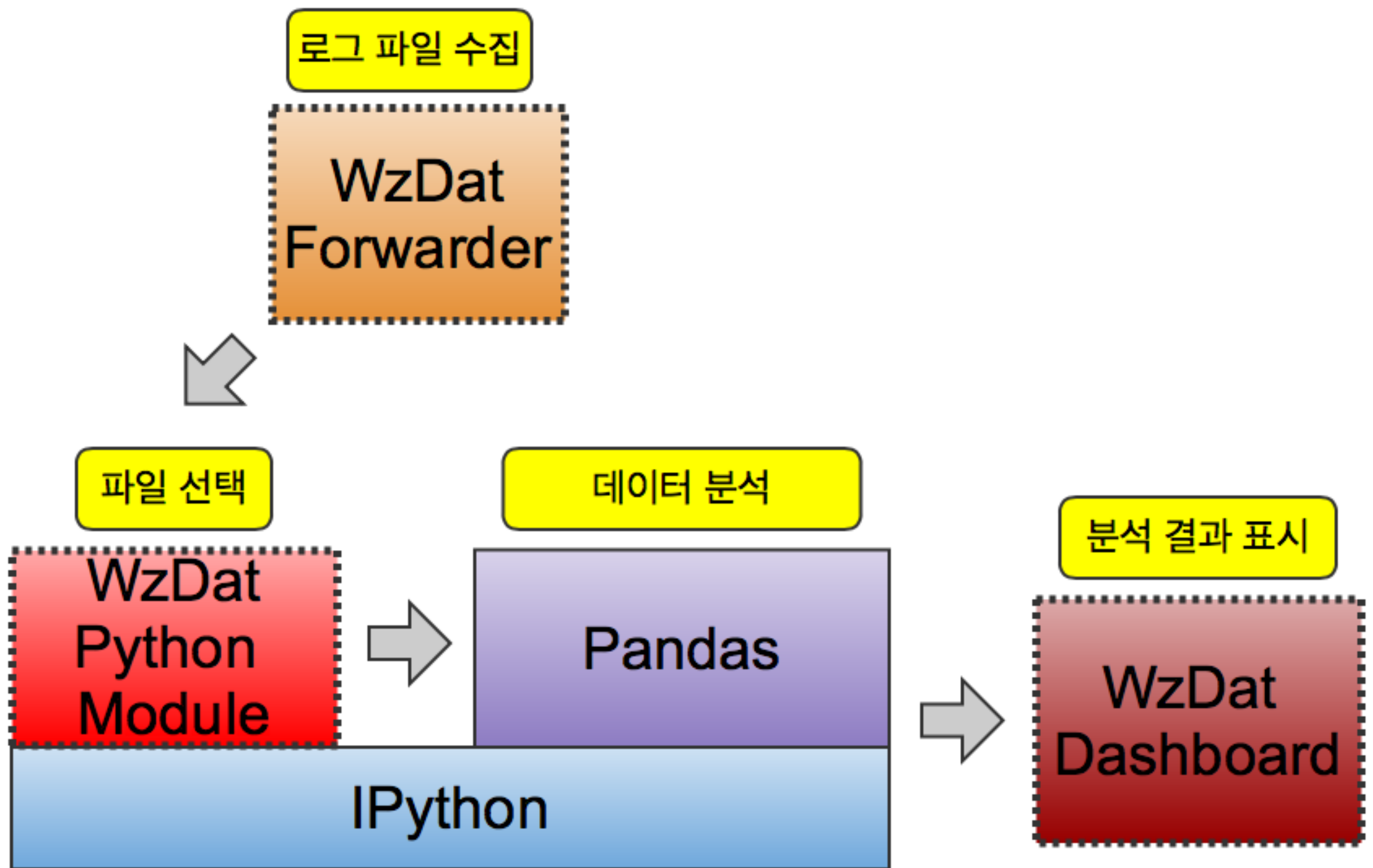
# 비전

- 오래된 프로젝트에서 코드 수정없이 쓸 수 있게 하자.
- 누구나 Raw 데이터에 접근할 수 있게 하자.
- 분석 과정과 결과를 다른 사람들과 공유하자.
- 분석을 자동화하고, 시각화하자



# 구성

- WzDat 포워더
- WzDat 파이썬 모듈
- WzDat 대쉬보드



# 포워더(Forwarder)

- 파이썬으로 만든 윈도우 서비스 (py2exe)
- DB, 로그파일, 시스템 로그(예정)등을 게임/DB 서버에서 분석 서버로 전송
- 설정파일(.yaml)에 동기화 정책 기술



# WzDat 파이썬 모듈

- 많은 로그파일 중 원하는 것을 빠르게 찾아 냄
  - 파일 타입 - log, dblog, dmp
  - 셀렉터 - files, nodes, servers, dates
  - 상수 필터링 - node, server, date
- 내용 검색 - find
- Pandas 데이터 프레임으로 변환 - to\_frame

# WzDat 파이썬 모듈 시연

[youtu.be/rygXJ4G66Ew](https://youtu.be/rygXJ4G66Ew)

# 대쉬보드(Dashboard)

- IPython 노트북에서 분석된 결과만 표시
- 파일 명으로 간단히 스케줄링 및 그룹짓기
  - ex) “[0 5 \* \* \* @중요] 데일리지표” - 매일 5시에 분석 실행 후 ‘중요’ 그룹에 ‘데일리지표’ 항목으로 등록
- UI컨트롤을 사용하여 대량의 데이터 중 일부만 선택해 보기



# WzDat 대쉬보드 시연

[youtu.be/1mdiiGmZTzc](https://youtu.be/1mdiiGmZTzc)

후기

# 개발하며 느낀점

- TDD
  - 인터프리터 언어는 테스트가 중요
- 문서화
  - 만든 사람도 까먹는다. 나를 위해서 시작하자.
- Git(hub)을 잘 활용하자.
  - VCS를 잘 쓰는 것은 프로그래밍 능력과 이어짐
- 윈도우용 배포를 위해서 py2exe 쓸만하다.



# 데이터 분석하며 느낀 점

- 현장의 목소리에 귀기울여야..
  - 도메인 지식이 없으면 무엇을 봐야할 지 모른다.
- Pandas로 대용량 분석을 위해서는 HDF를 잘 활용하자
  - 빅데이터는 아니어도 HDD 용량만큼 든든
- Pandas 쿼리 최적화 공부 필요
  - 경험이 쌓일 수록 쿼리가 짧아진다. 명료하고 빨라질 가능성
- 사용자 친화적인 시각화를 위해 고민하자.

# 배포하며 느낀점

- 플랫폼으로 제공하려니 설치가 너무 번거롭다.
  - IPython + Pandas, uWSGI, nginx, Celery, Redis, SQLite3, HDF5, Supervisor..
  - 소스 clone 후 pip로 의존 모듈 설치, 각종 스크립트 및 Crontab 설정, 방화벽 등 시스템 설정..

- Docker!



- `docker pull hajee01/wzdat`

# 도움주신 분들

- 스마트스터디 박현우 이사님 (WBERS!)
- 웹젠애플레이 이인규 대표님
- 파이썬 코리아 강남 스터디 여러분들



들어주셔서 감사합니다.👍

게임 데이터 분석에 관심있으신 분은 연락주세요~

[haje01@naver.com](mailto:haje01@naver.com)