

파이썬, R 기반 빅데이터 플랫폼 개발자 ; 빅데이터 시각화 구현

- 이 소 영 강사

yisy0703@naver.com

<https://github.com/yisy0703/bigdata>

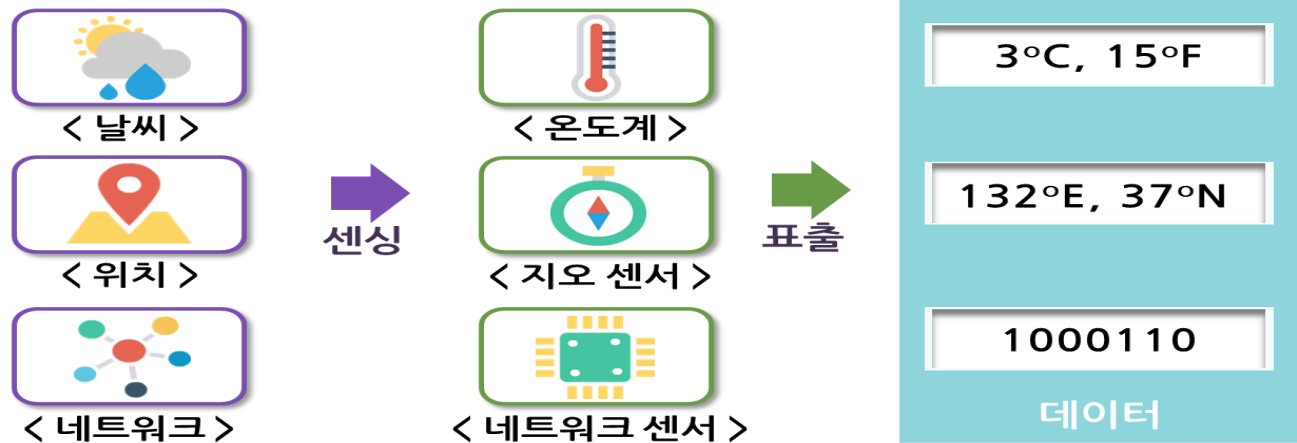
₩₩192.168.1.165

빅데이터 처리 기술의 필요성

1) 인간과 데이터



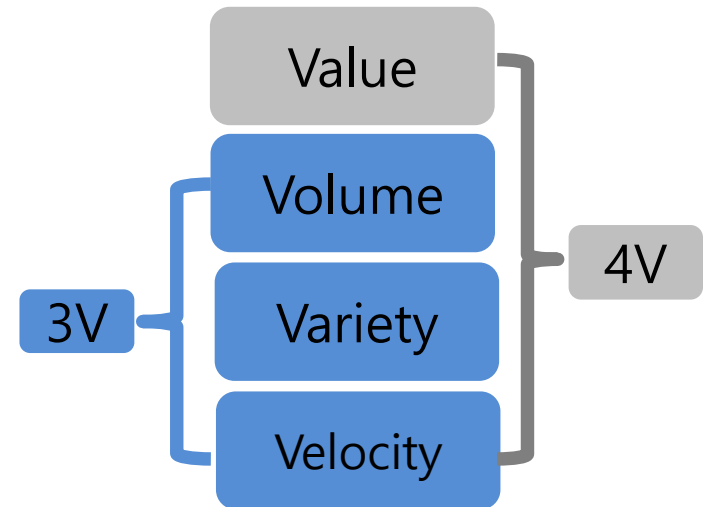
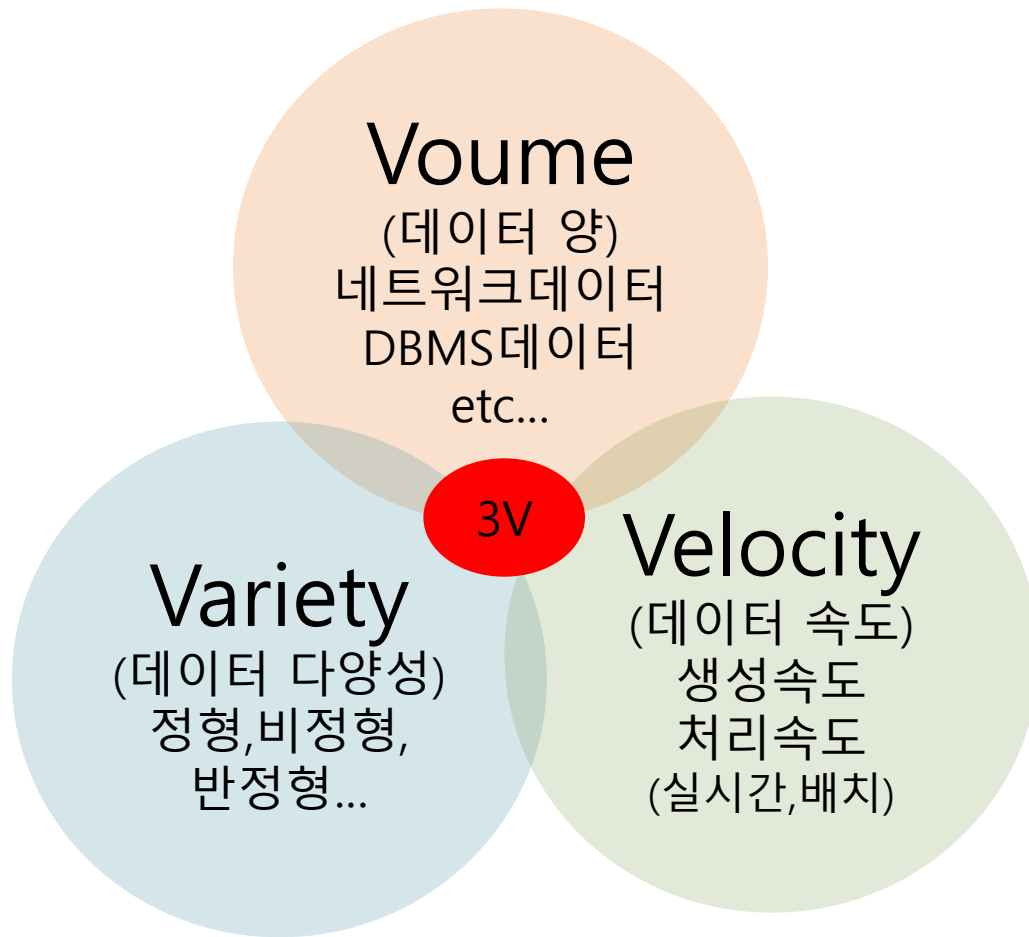
2) 센서와 데이터



빅데이터의 특성

킬로(10^3)-메가(10^6)-기가(10^9)-테라(10^{12})-페타(10^{15})-엑사(10^{18})-제타(10^{21})-요타(10^{24})

- 빅데이터의 특성 : 3V + Value
 - ✓ 초기 가트너에서 정의한 빅데이터의 3V에 가치를 더한 개념



빅데이터의 속성 정리

빅데이터
소스유형

활용방향

데이터 형태

저장 소스

① 휴먼 데이터

(ERP그룹웨어, CRM고객데이터, MES 제조시스템, SCM물류시스템, VOC고객센터 데이터...)

② 로그성시스템데이터

(기계 ex. 서버, 방화벽, 건물(bams), 중장비, 차량센서, IoT, 스마트폰 등으로부터 나온 데이터)

③ 웹 데이터

(기사 및 sns 그 외 다양한 웹 데이터)

● 현황파악(관계분석, 문맥데이터분석의 자연어 처리)

— 행태분석을 통한 타켓 마케팅, 미래예측, 재난 안전, 이탈고객방지 ..

+ 신상품개발, 고객유치, (분석과 전략이 함께)

기준 :

- ① 스키마가 유무?
- ② 연산가능 ?

정형데이터

반정형데이터
(로그 text, xml, json파일)

비정형데이터
(text, image, voice, video)

RDBMS
(Oracle, MySQL)

txt, xml, json
파일형태

하둡기반 HBASE
NoSQL

비정형데이터는 정형화 어플리케이션의 성능이 중요(java, R, Python)



RDBMS

Sqoop

HDFS

Spark
(Spark, Java, Python, R, SQL)

IoT기기
로그데이터

NoSQL

MapReduce
(Java)

공공데이터
(data.go.kr.)

RDBMS

Pig

HIVE

Web데이터
(기사, SNS...)

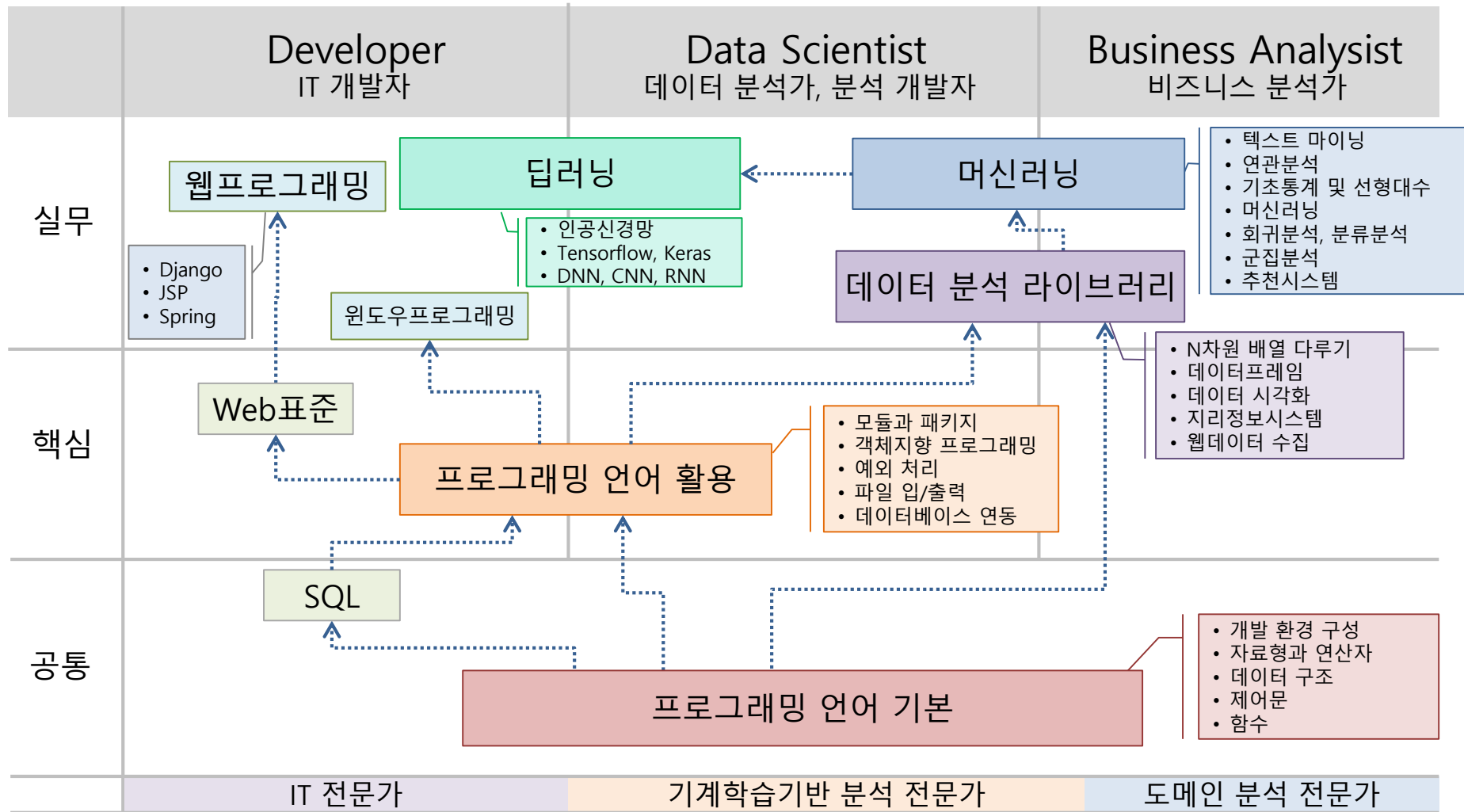
크롤러
(Python, R, Java)

R, Python
(머신러닝 - 통계기반,
딥러닝-지도학습과 비지도학습)

일정

- ✓ JAVA 4주
- ✓ DB (oracle, MySQL) 2주
- ✓ 웹표준(HTML, CSS, JavaScript, jQuery) 2주
- ✓ R 3주
- ✓ Python 3주
- ✓ 머신러닝, 딥러닝, 자연어처리 1주
- ✓ dJango 1주
- ✓ Hadoop 에코 시스템 1주
- ✓ 프로젝트(개별프로젝트, 팀프로젝트) 3주
- ✓ 취업지원팀 : 이력서 특강, 취업특강 등

로드맵



교과목 (NCS.go.kr참조)

	NCS능력단위	분류번호	시간	서술형	포트폴리오	사례연구	평가일자	평가일시
1	개발자 환경 구축	2001020233_19v4	40	○	○		5월17일	6일차
2	프로그래밍 언어 활용	2001020231_19v4	50	○	○		5월26일	12일차
3	프로그래밍 언어 응용	2001020230_19v4	50	○	○		6월3일	18일차
4	데이터 입출력 구현	2001020205_19v5	40	○	○		6월10일	23일차
5	응용SW 기초기술 활용	2001020232_19v4	40	○	○		6월17일	28일차
6	화면 설계	2001020224_19v5	40		○		7월1일	38일차
7	화면 구현	2001020225_19v5	40		○		7월1일	38일차
8	GUI 상세디자인	2001020711_19v3	60		○		7월13일	46일차
9	빅데이터 분석 결과 시각화	2001010509_19v3	40		○	○	7월20일	51일차
10	UI 디자인	2001020707_19v3	105		○		8월6일	64일차
11	UI 제작	2001020712_19v3	60		○		8월17일	71일차
12	UI 구현	2001020708_19v3	100		○		9월3일	84일차
13	빅데이터 수집시스템 개발	2001020903_19v2	40	○	○		9월15일	92일차
14	분석용 데이터 구축	2001010510_19v3	20		○	○	9월15일	92일차
15	애플리케이션 테스트 수행	2001020227_19v5	30	○	○		9월23일	95일차
		집체 수업 시간 합	755					
원격학습	의사소통능력[원격]	5/28 ~ 6/03	10					
원격학습	UI 디자인[원격]	6/04 ~ 6/11	16					
원격학습	화면 구현 part1[원격]	6/16 ~ 6/22	12					
원격학습	화면 구현 part2[원격]	6/25 ~ 7/01	12					
원격학습	빅데이터 분석 결과 시각화[원격]	7/07 ~ 7/21	20					
원격학습	빅데이터 수집시스템 개발[원격]	9/06 ~ 9/16	15					
		온라인 수업 시간 합	85					

시스템 폴더 관리

d:\WbigData\DOWNLOAD : 개발환경을 위한 다운로드 파일

d:\WbigData\WIDE : 통합개발환경 폴더

d:\WbigData\Wsrc : 수업시간 개발 소스

d:\WbigData\Wnote : 교안 노트

d:\WbigData\W취업 : 취업에 관련된 정보 및 이력서 포트폴리

오(버전 관리 중요)

시스템 설치

1. Vaccine(V3) 다운로드 & 설치 <https://www.ahnlab.com/>
2. Chrome 브라우저 <https://www.google.com/chrome/browser>
3. 자바 : <https://www.oracle.com/>
4. 이클립스 IDE(java 8이 가능한 eclipse 4.16 download) :
<http://archive.eclipse.org/eclipse/downloads/drops4/R-4.16-202006040540/> (참조. <https://yyman.tistory.com/1451?category=316363>)
5. Git <https://git-scm.com>

깃허브의 레파지터리 생성

1. <http://www.github.com>에 가입한 후 확인지침이 포함된

메일 확인 후 start a Project

2. git-scm.com 사이트에 다운받은 파일을 인스톨

➤ `git config --global user.name "가입한 사용자명"`

➤ `git config --global user.email "가입한 이메일 주소"`

➤ README.md를 작성하고 cmd에서 새 저장소를 작성

깃허브의 레파지터리 생성(main)

- `git init`
- `git add README.md`
- `git commit -m "first commit"`
- `git branch -M main`
- `git remote add origin git경로`(git start project 페이지에 명시 예.

`https://github.com/yisy0703/xx.git`)
- `git push -u origin main`

두번째 업로드부터(main)

- git init
- git add 폴더명이나 파일명
- git commit -m "05-10"
- git remote add origin git경로(git start project 페이지에 명시 예.
<https://github.com/yisy0703/xx.git>)
- git push -u origin main

깃허브의 레파지터리 생성(master)

- `git init`
- `git add README.md`
- `git commit -m "first commit"`
- `git remote add origin git경로(git start project 페이지에 명시 예.
https://github.com/yisy0703/xx.git)`
- `git push -u origin master`

두번째 업로드부터(master)

- git init
- git add 폴더명이나 파일명
- git commit -m "05-10"
- git remote add origin git경로(git start project 페이지에 명시 예.
<https://github.com/yisy0703/xx.git>)
- git push -u origin master