

Python 기반 머신러닝/딥러닝 실습

(1일차 – 4시간)

2022. 6.

정 준 수 Ph.D

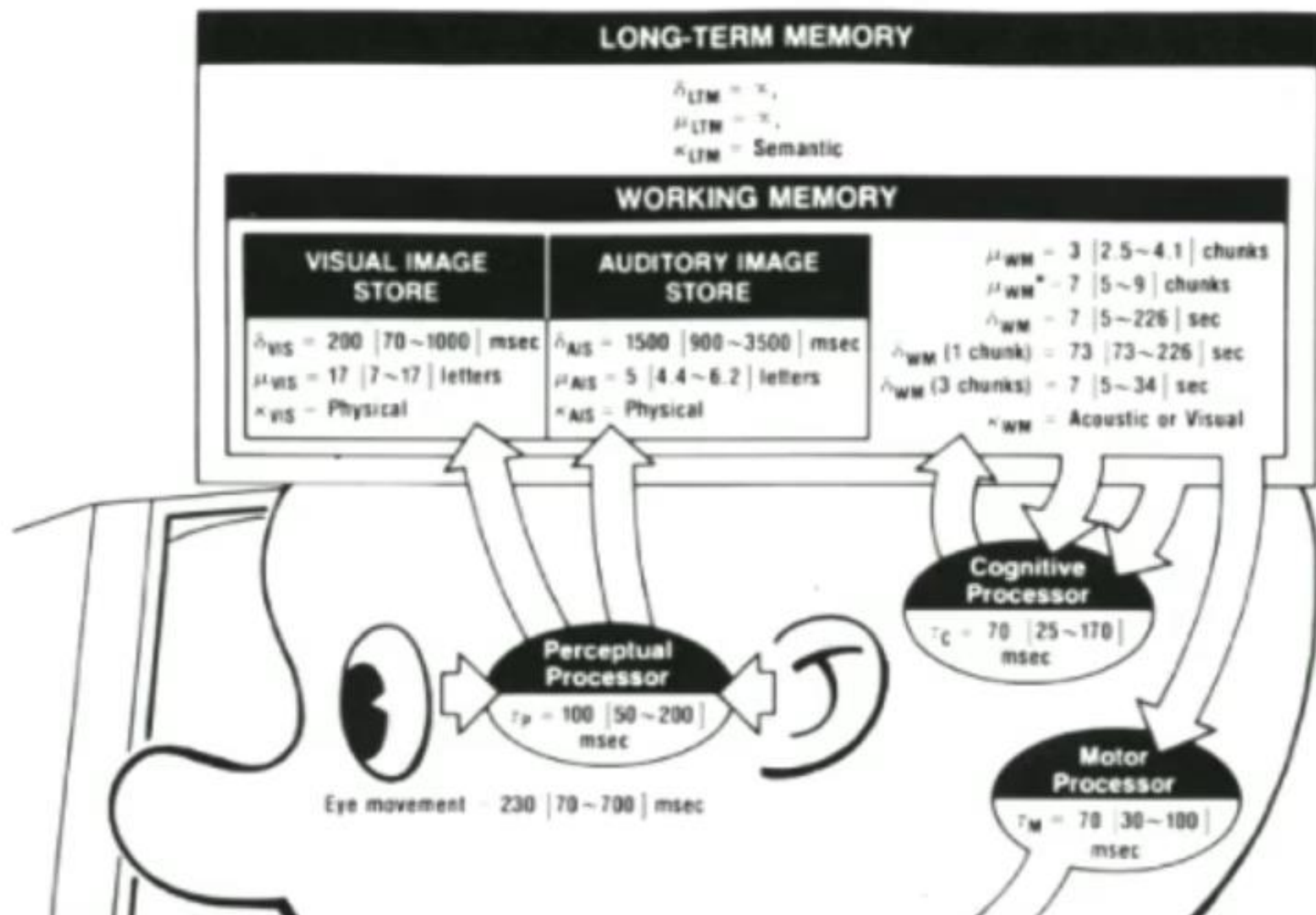
과정 목표

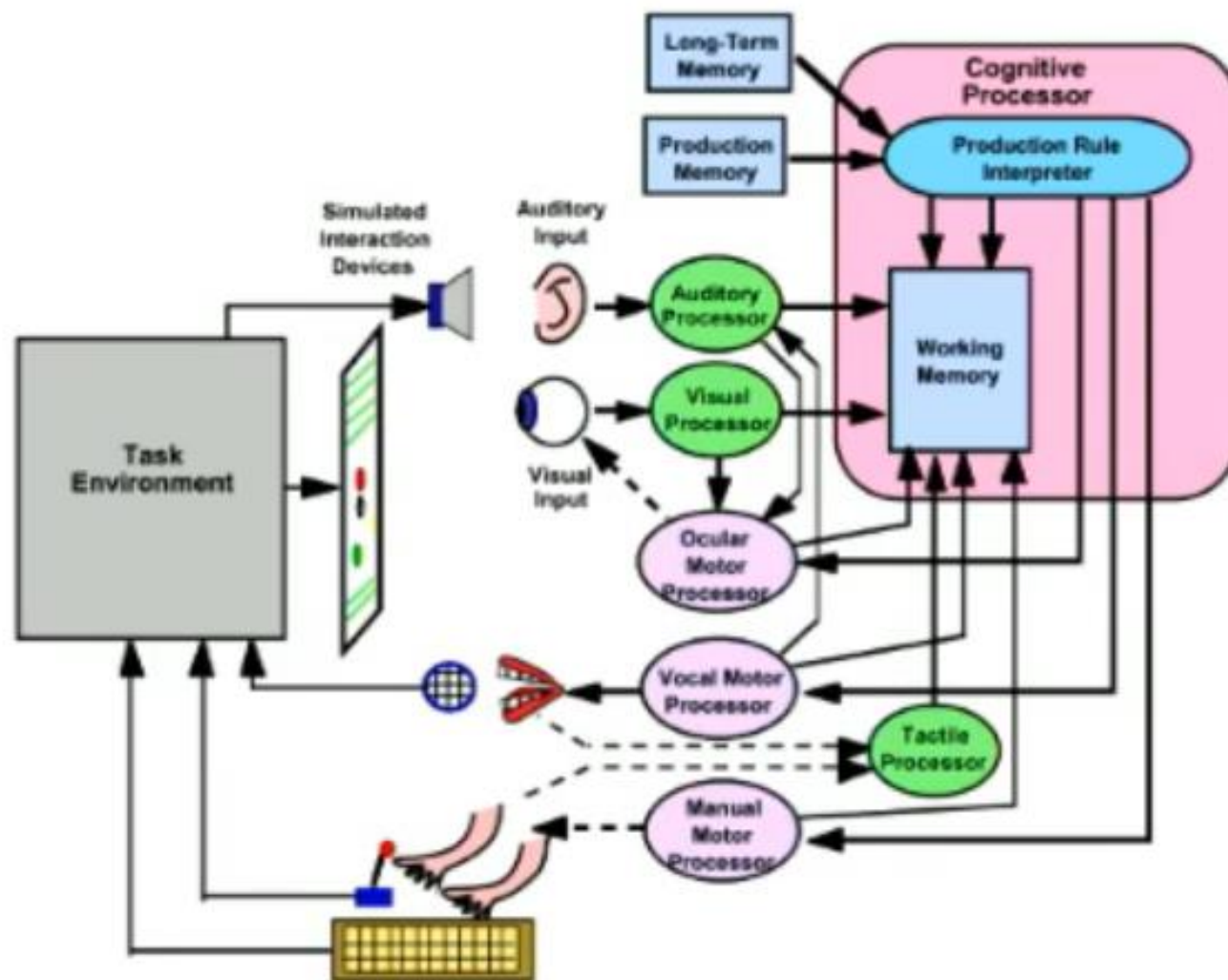
빅 데이터 분석 프로젝트 수행을 위한 데이터 분석 기획, 수집, 분석
및 보고서 작성에 필요한 내용 학습

1. Python 활용(객체지향 파이썬 프로그래밍)
2. 데이터 처리 패키지를 활용한 통계 분석
3. 분석 방법론, 빅데이터 분석, 머신러닝을 활용한 예측분석(Predictive Analytics)

Artificial intelligence started as a field whose goal was to replicate human level intelligence in a machine.

인공 지능은 인간 수준의 지능을 기계에 복제하는 것을 목표로 시작되었습니다.





The EPIC architecture
for cognition
and performance
with application to
human-computer interaction

Software란?

컴퓨터 프로그램이란?

알고리즘의 추상화

★ 생각해 보기

질문) 프로그램이 먼저 나왔을까요? 컴퓨터가 먼저 나왔을까요?

객체 지향 프로그래밍 object-oriented programming (OOP)

OOP focuses on the objects that developers want to manipulate rather than the logic required to manipulate them.

Class(클래스)

객체를 정의해 놓은 것
객체를 정의하는 틀 또는 설계도





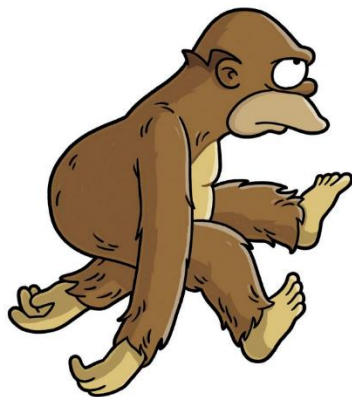
Object(객체)

클래스의 인스턴스
설계도를 통해 만들어진 것

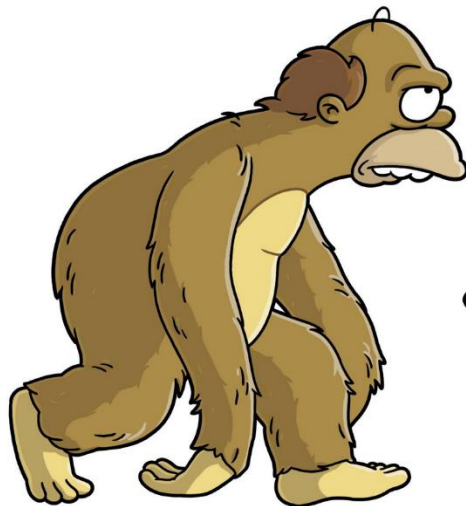
함수형 언어(Functional program)란?



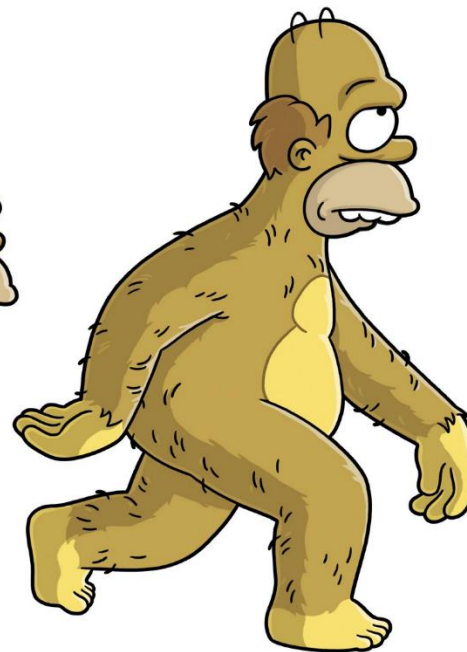
MACHINE



ASSEMBLY



PROCEDURAL



OBJECT ORIENTED



FUNCTIONAL

MATT GROENING

함수형 프로그래밍을 지원하는 언어들

Support Functional Programming



JavaScript



Kotlin



Go



Python



Rust



Java



C++



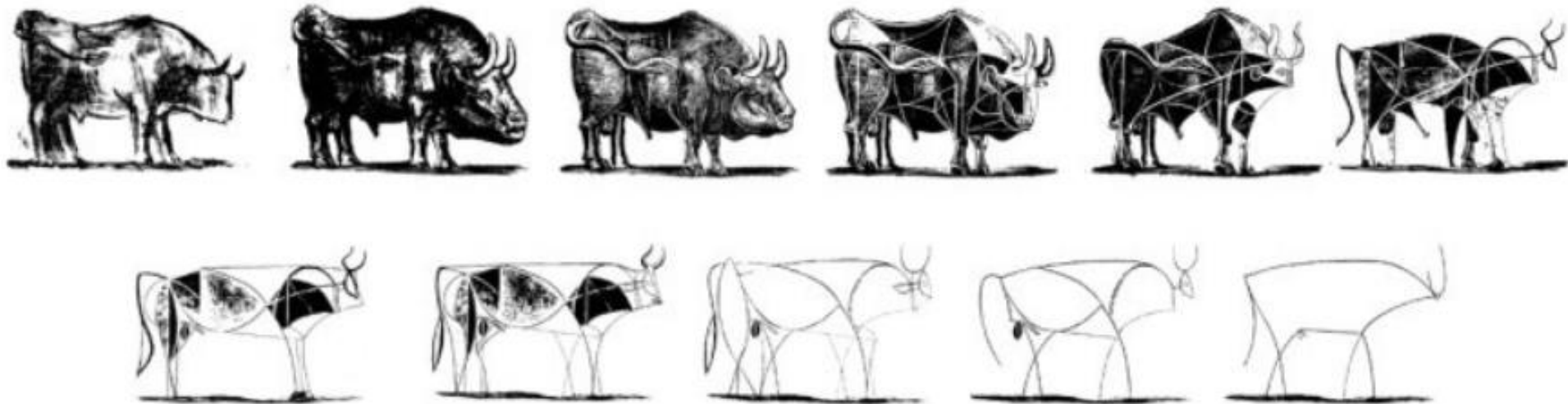
Scala

Why Python?

추상화(Abstraction)

컴퓨터 과학에서 추상화(abstraction)는 복잡한 자료, 모듈, 시스템 등으로부터 핵심적인 개념 또는 기능을 간추려 내는 것을 말한다.

The Evolution of Picasso Bull



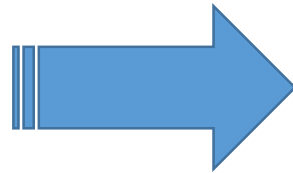




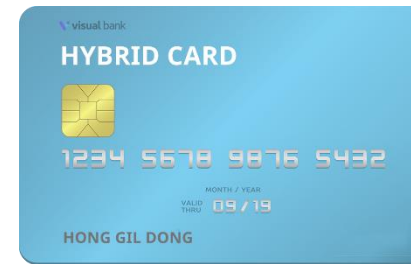
현금(Money)의 추상화(Abstraction) 사례



현금(Money)



인터넷 뱅킹



신용카드



티머니

현금(Money)의 추상화(Abstraction) – Featuring 해보기



현금과 다른 특징 분석

추상적 자료형(abstract data type, ADT)

컴퓨터 과학에서 자료들과 그 자료들에 대한 연산들을 명기한 것이다

1. 구조(Structure)
2. 연산(Operation)

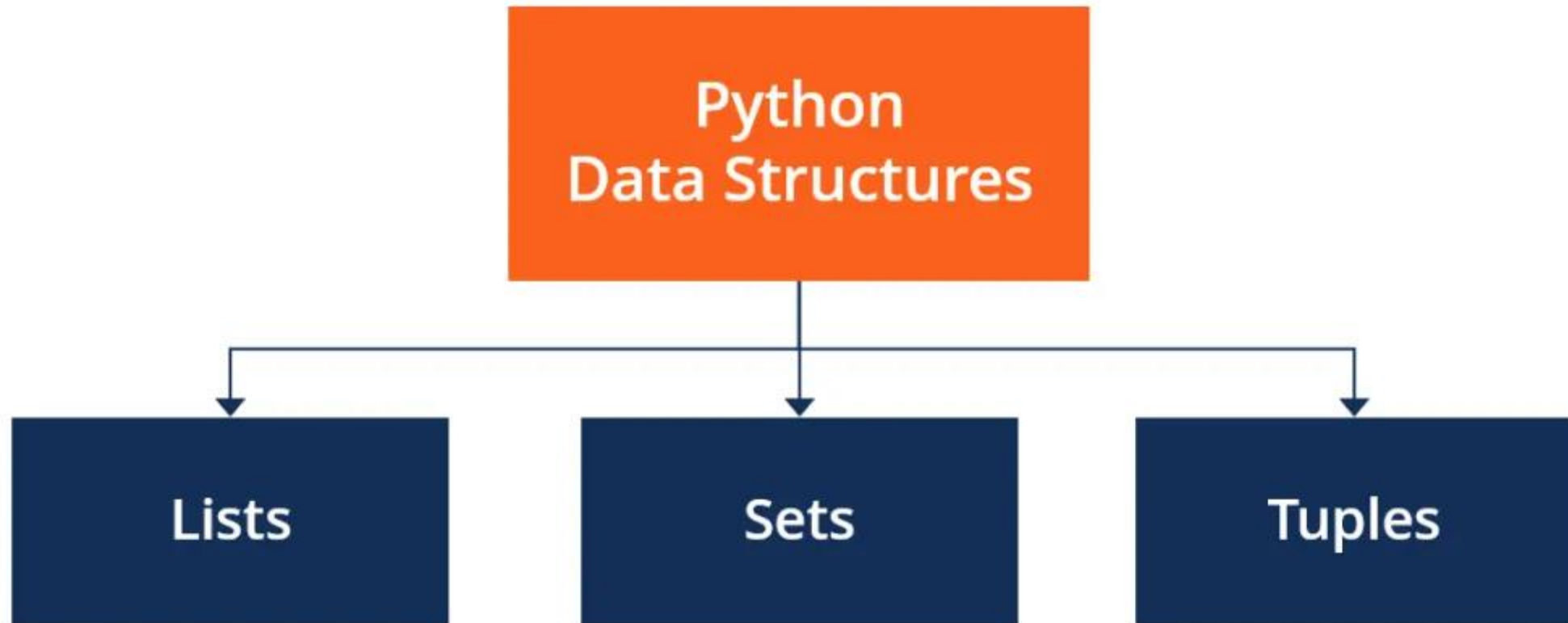
Python 자료형이란?



- ◉ 일반적으로 프로그램은 연산을 수행하며, 연산을 하기 위해서는 값들을 임시로 혹은 영구적으로 어딘가에 저장해 놔야 함
- ◉ 이러한 값들을 저장하는 저장소를 변수라고 함
- ◉ 변수의 유형을 자료형이라고 함

Python 자료의 종류

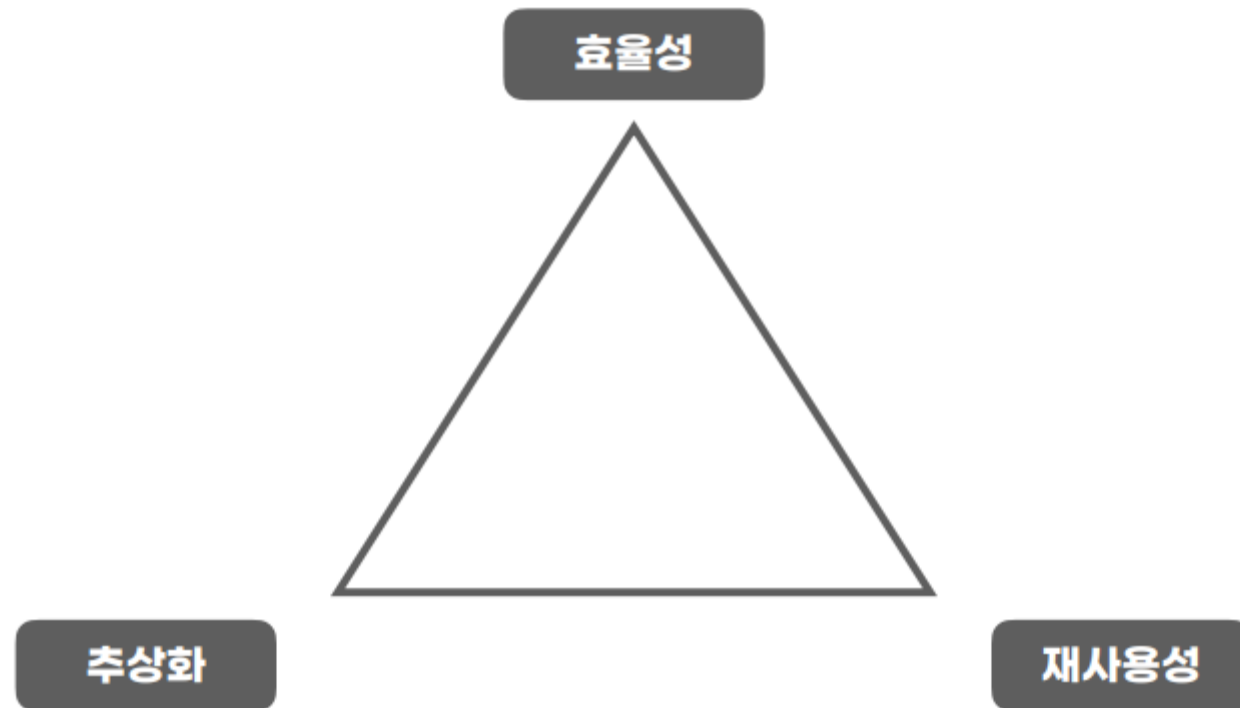
- 수치 - int float complex 등
- 문자형
- 리스트 - 리스트는 쉽게 값들의 나열
- 세트 - 집합의 개념으로 순서가 없음
- 튜플 - tuple은 리스트와 유사하나, 읽기 전용임
- 딕셔너리 - 키와 값의 쌍으로 이루어져 있음
- 부울(bool) - 참(True)과 거짓(False)을 나타내는 자료형



<실습 예제 프로그램>

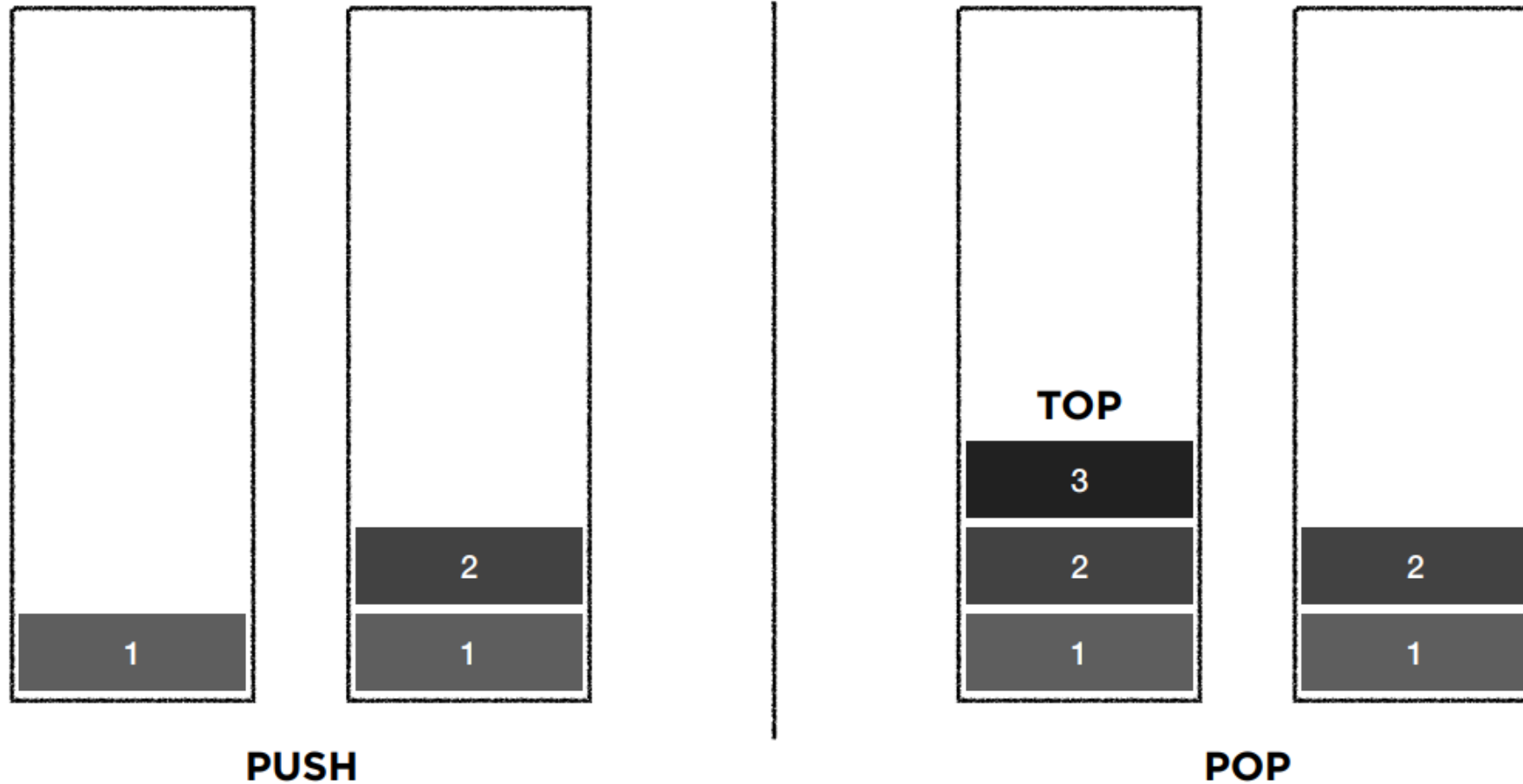
https://github.com/JSJeong-me/Machine_Learning/tree/main/Python_Data_Types

자료구조

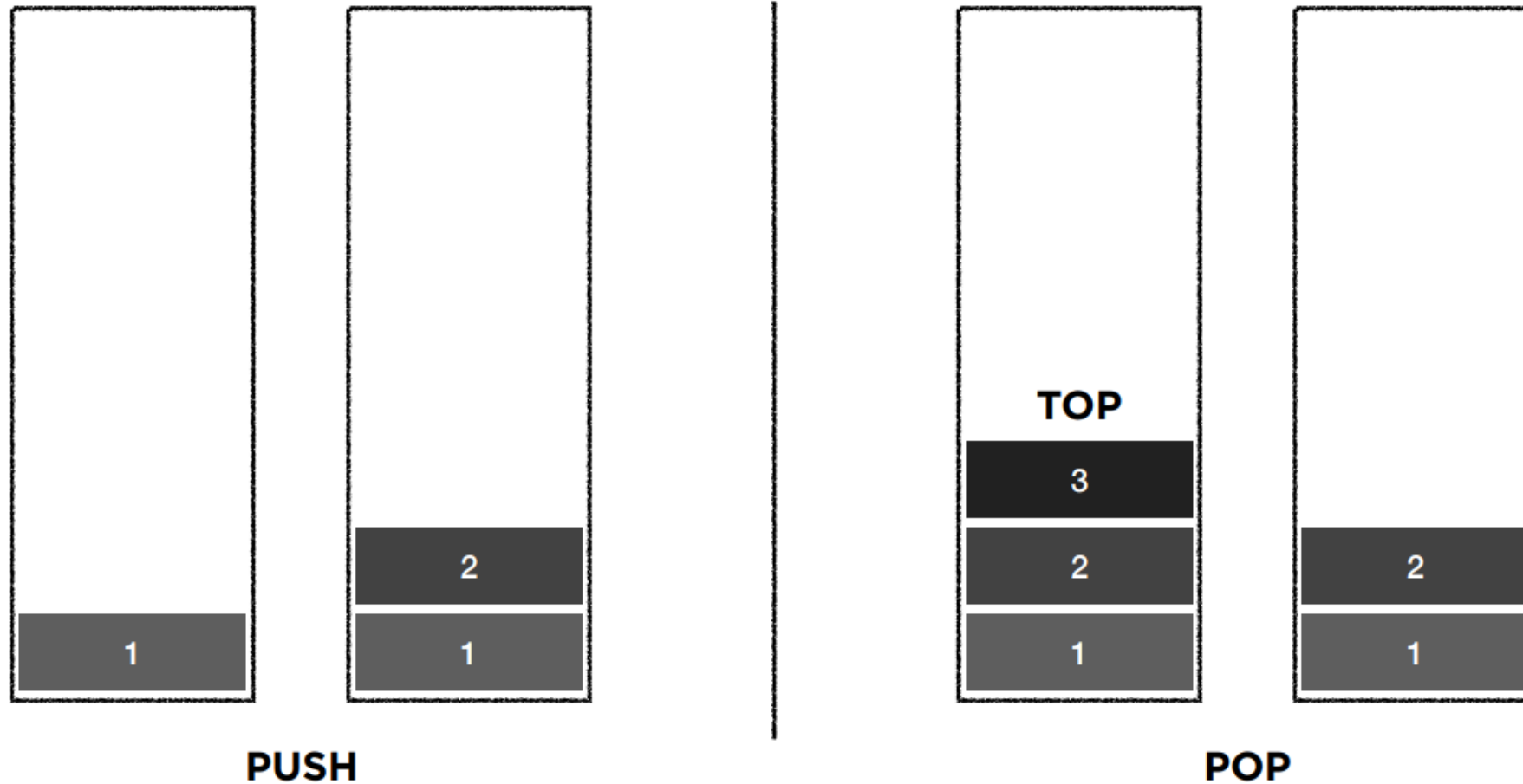


자료를 저장하는 방법론, 규칙

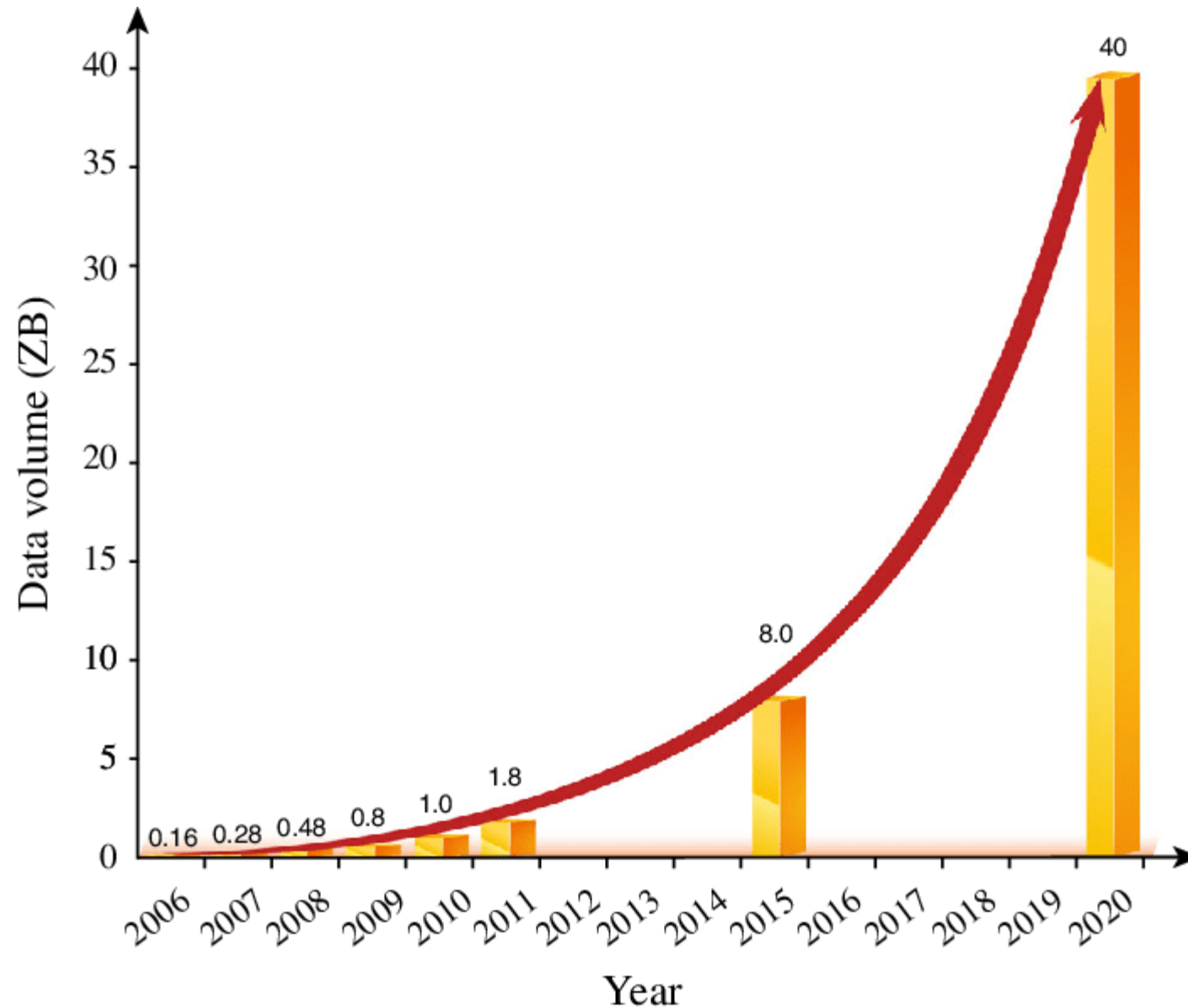
Stack: FILO(First In, Last Out)



Queue: FIFO(First In, Fast Out)



Big Data



Global growth trend of data volume, 2006-2020 (based on "The digital universe in 2020: big data, bigger digital shadows, and biggest growth in the far east")

Machine Data

Log parsing

Raw Log

Mar 19 2020 23:53:23:
%ASA-4-338002: Dynamic Filter
monitored blacklisted UDP traffic
from
Sample_Outside:238.134.165.47/
39266 (238.134.165.47/39266) to
Sample_LinuxDB:238.134.165.47/
38274

Log Parser

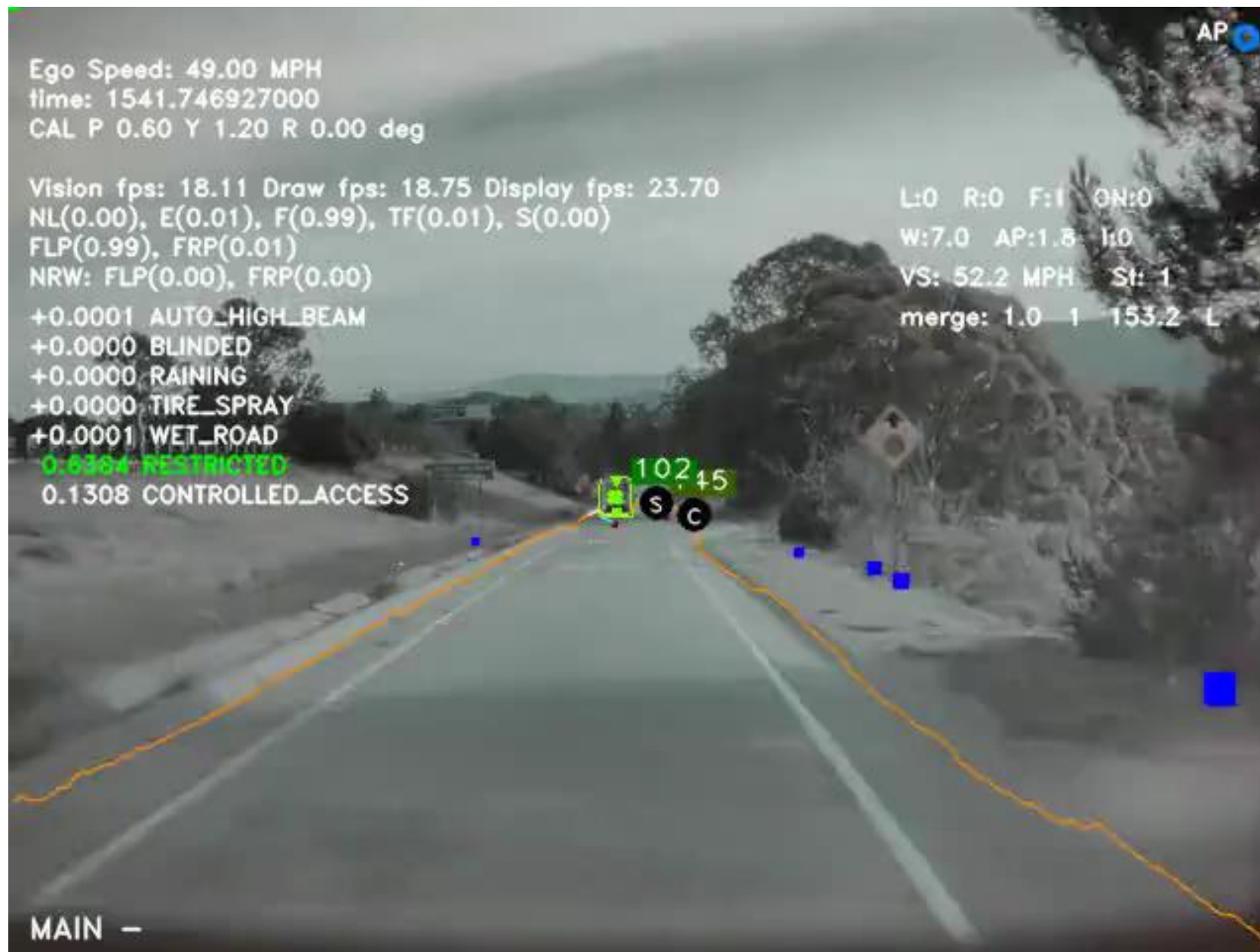


Parsed Log

Field names	Field values
timestamp	Mar 19 2020 23:53:23
product	ASA
log_level	4
message_id	338002
source_zone	Sample_Outside
source_ip	238.134.165.47
source_translated_ip	238.134.165.47
dest_zone	Sample_LinuxDB
dest_ip	238.134.165.47
dest_translated_ip	238.134.165.47/38274

Downstream





강사 소개

정 준 수 / Ph.D (jsjeong@hansung.ac.kr)

- 前) 삼성전자 연구원
- 前) 삼성의료원 (삼성생명과학연구소)
- 前) 삼성SDS (정보기술연구소)
- 現) (사)한국인공지능협회, AI, 머신러닝 강의
- 現) 한국소프트웨어산업협회, AI, 머신러닝 강의
- 現) 서울디지털재단, AI 자문위원
- 現) 한성대학교 교수(겸)
- 전문분야: 시각 모델링, 머신러닝(ML), RPA
- <https://github.com/JSJeong-me/>

