# 인공지능이해하기

## 정리하기

# 변화 시기 이해하기

- 빅데이터의 크기는 고정되지 않는다
- (사례)빅데이터의 사용은 융합의 형태로 다양하게 사용된다
- 데이터가 대량으로 생산되게된 이유는 무엇인가 ?
  - 데이터 생산에 사물이 참여
  - 비정형 데이터의 증가
- 빅데이터 처리를 위해서는 고속데이터 처리 하드웨어와 클라우드가 필요
- 빅데이터 처리에 파이썬이 사용되는 이유는
  - 다양한 데이터 분석을 위한 라이브러리를 제공

# 변화의 시대

## 인공지능 활용과 제약점

- 기계학습 딥러닝은 마술이 아님
- 데이터를 다루는 수학과 통계 모델

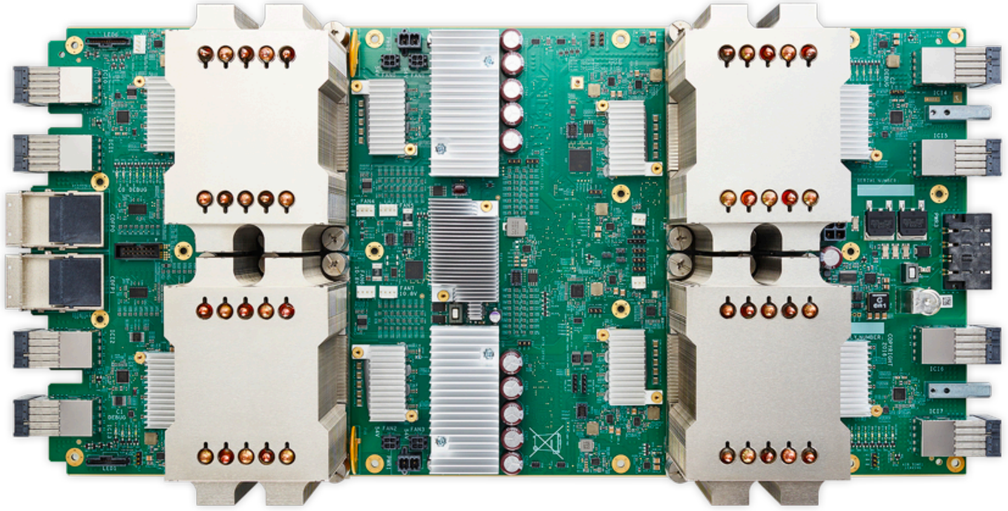
# *빅데이터

## 처리과정과 기술

# *빅데이터

## AI 프로세서

- Tensor Processing Unit (TPU)
  - 기계학습 특화장치



# 데이터과학 활용
## 의료

- Big Data + Computing = Improved Health?

### Traditional Model

**1. Hypothesis:** An early study suggests that patients with Gaucher's disease (caused by a mutation to the GBA gene) might be at increased risk of Parkinson's.
**2. Studies:** Researchers conduct further studies, with varying statistical significance.
**3. Data aggregation**: Sixteen centers pool information on more than 5,500 Parkinson's patients.
**4. Analysis:** A statistician crunches the numbers.
**5. Writing:** A paper is drafted and approved by 64 authors.
**6. Submission:** The paper is submitted to *The New England Journal of Medicine*. Peer review ensues.
**7. Acceptance:** *NEJM* accepts the paper.
**8. Publication:** The paper notes that people with Parkinson's are 5.4 times more likely to carry the GBA mutation.

**Total time elapsed: 6 years**

### Parkinson's Genetics initiative

**1. Tool Construction:** Survey designers build the questionnaire that patients will use to report symptoms.
**2. Recruitment:** The community is announced, with a goal of recruiting 10,000 subjects with Parkinson's.
**3. Data aggregation:** Community members get their DNA analyzed. They also fill out surveys.
**4. Analysis:** Reacting to the *NEJM* paper, 23andMe researchers run a database query based on 3,200 subjects. The results are returned in 20 minutes.
**5. Presentation:** The results are reported at a Royal Society of Medicine meeting in London: People with GBA are 5 times more likely to have Parkinson's, which is squarely in line with the *NEJM* paper. The finding will possibly be published at a later date.

**Total time elapsed: 8 months**

# 데이터과학 활용

## 심야 버스 노선

- 택시 승하차 정보, KT 통화량 데이터