
과학기술정보통신부
nipa 정보통신산업진흥원
National IT Industry Promotion Agency
2021
공개SW
페스티벌
ONLINE FESTIVAL

2021 공개SW 페스티벌
ONLINE FESTIVAL

개발자와 함께 성장하는 오픈소스

# 오픈소스로 보는 데이터 생태계의 발전

소속 : Harmonize
이름 : 한기용

과학기술정보통신부 nipa 정보통신산업진흥원 National IT Industry Promotion Agency

# 목차

# 연사소개


YAHOO! SEARCH

polyvore™

udemy

SAMSUNG ELECTRONICS

# 80-90년대의 인공 신경망

- 90년대 초반 대학원 전공으로 고려
- 하지만 다양한 문제 존재
  - 컴퓨팅 파워 부족이 그 중 하나

# 검색엔진 개발 #1

- 구글 페이지랭크 논문 발표 (1998)
- 이후 웹구조 기반 페이지 중요도 계산이 일반화
  - 단 이는 대용량 데이터 처리를 필요로 함

# 검색엔진 개발 #2

- 좋은 빅데이터 분산 처리 시스템이 존재하지 않았음
  - 야후 검색: 크롤/웹맵/인덱싱/검색 시스템이 개별 구현됨
  - 중복 개발과 관리로 인한 이슈 발생
- 구글 맵리듀스('04)와 빅테이블('06)논문 발표
  - 하둡 오픈소스 개발의 기반이 됨

# 검색엔진 개발과 오픈소스의 만남

- Doug Cutting
- Lucene/Nutch -> Hadoop (HDFS/MapReduce)
- 2006년부터 야후에서 하둡을 풀타임으로 개발

# 하둡의 파급 효과

- 생태계를 만들어냄
  - Hive/Presto, Oozie, …
  - YARN, Spark, …
- IT 기업들의 오픈소스 개발
  - 야후, 넷플릭스, Airbnb, …
- 오픈소스를 상업화하는 기업들의 탄생
  - Cloudera, HortonWorks, MapR, …

# 머신러닝과 인공지능의 각광

- 빅데이터 처리가 가능해지고 클라우드를 기반으로 대용량 컴퓨팅 환경의 사용이 전보다 쉬워지면서 딥러닝이 다시 부상
- Tensorflow, PyTorch, MXNet등의 오픈소스 기반 머신러닝 프레임웍이 배포되기 시작

# 맺음말

- 개인적인 경험을 바탕으로 구글과 오픈소스가 데이터 분야에 끼친 공헌에 대해 살펴봄
- 오픈소스는 선택이 아니라 필수

"Open source is a requirement for business."
- Doug Cutting

# 감사합니다

과학기술정보통신부

nipa 정보통신산업진흥원
National IT Industry Promotion Agency