

# BigData for Cloud (PaaS-TA)

INNOVATION & FOCUS



# DATA?

## 정형 데이터 vs 비정형 데이터

정형 데이터 ?  
일정한 규격과 형태를 가진 데이터 형식

비정형 데이터 ?  
그림, 영상, 문서 형태와 구조가 각기 다른 구조화 되지 않은 데이터 형식



# SOCIAL NETWORK SERVICE

빅데이터의 85% 가량이 비정형 데이터  
급속도로 확산되고 있는 추세  
정형 데이터 분석을 위해서 이용되고 있는 많은  
기술들이 비정형 데이터에서는 활용할 수  
없다라는 한계

이러한 이유로 하둡이 이슈



INNOCUS

# 빅데이터 분석기법 1/2

데이터 양이 방대하고 분석해야 할 비정형 데이터 비중이 높아서 정확한 정보 추출 기술이 중요

텍스트마이닝 (Text Mining)

텍스트 마이닝은 비/반정형 텍스트 데이터에서 자연어처리(Natural Language Processing)기술에 기반하여 유용한 정보를 추출, 가공하는 것을 목적으로 하는 기술



문서분류(Document Classification)



문서군집(Document Clustering)



정보추출(Information Extraction)



문서요약(Document Summarization)



INNOCUS





## 빅데이터 분석 기법 2/2

### 오피니언 마이닝 (Opinion Mining)

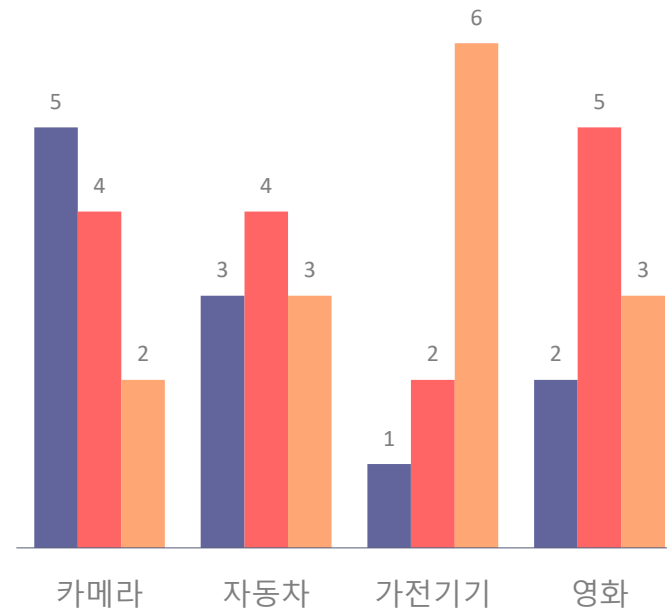
텍스트 마이닝의 관련 분야로는 오피니언 마이닝, 혹은  
평판분석(sentiment Analysis)이라고 불리는 기술이 있고 오피니언  
마이닝은 소셜미디어 등의 정형/비정형 텍스트의 긍정(Positive),  
부정(Negative), 중립(Neutral)의 선호도를 판별하는 기술

### 소셜 네트워크 분석 (Social network Analytics)

소셜 네트워크 분석은 수학의 그래프 이론(Graph Theory)에 뿌리를  
두고 소셜 네트워크 연결구조 및 연결강도 등을 바탕으로 사용자의  
명성 및 영향력을 측정하여 소셜 네트워크 상에서 입소문의  
중심이나 허브(Hub) 역할을 하는 사요자를 찾는데 주로 활용

### 클러스터 분석 (Cluster Analysis)

군집분석은 비슷한 특성을 가진 개체를 합쳐가면서 최종적으로  
유사 특성의 군집(Group)을 발굴하는데 사용 ex) 트위터상에서  
사진/카메라 대해 이야기하는 사용자집단 자동차에 관심있는  
사용자집단을 군집분석을 통해 분류



INNOCUS



“

클라우드 기반에서의 빅데이터는 처리는?

빅데이터 처리 인프라 기술.

”

**HPC?**

스마트 시티, 스마트 팩토리에 슈퍼컴퓨팅  
파워가 지원하는지?



“

**AI, 빅데이터와 CLOUD 연계는?**

IaaS, PaaS 인지 보안은 백업은 DR은



INNOCUS

# 빅데이터 처리 인프라 기술

Scale Out 대용량의 데이터를 분산 저장 HDFS

분산처리시스템 MapReduce, Spark...

Hadoop 의 방향은 ?

Hadoop1

- 클러스터당 최대 4000개의 노드를 등록가능
- MR 잡을 수행할 수 있음
- 작업처리를 슬롯 단위로 수행

Hadoop 2

- 클러스터당 10000개 이상의 노드가 가능
- MR 잡외에 Spark, Hama, Giraph 등 다른 분산 처리 모델도 수행가능
- Hadoop1의 잡트래커가 리소스 관리, 잡 스케줄링을 동시에 처리하여 병목 지점이 되어 Yarn 추가
- Yarn 은 리소스매니저, 어플리케이션 마스터가 리소스 관리, 잡스케줄링을 담당

Hadoop 3

- HDFS erasure coding 추가
- 기존의 Replication을 대체하는 방식으로 복제를 하지 않고 오류에 대응
- Yarn 타임라인 서비스 v2
- 기존 타임라인 서비스보다 많은 정보 확인 가능
- 헬스크립트 재작성
- 오래된 헬스크립트를 재작성하여 버그 수정
- JAVA8 지원

PS) 클라우데라랑 호튼웍스는 왜 인수합병 했을까?

클라우드 때문 !!! 그래서 CDP 가 탄생 ~!!

엣지부터 클라우드까지, 빅데이터에서 AI까지 분석의 분석을 지원





# HPC CLOUD

슈퍼컴퓨팅 파워와 클라우드.

## 필요성

빠른 속도의 데이터 증가속도와 실시간 처리

5G의 탄생 Massive IoT  
상상하지도 못할 속도로 증가되는 데이터  
실시간 분석이 필요한 세상



## Application, MSA

On HPC Cloud ? Or General Cloud

PaaS Platform 에서의 서비스  
PaaS 올리는 기반은 ???  
General Cloud 에서 Smart City 가 가능한가?

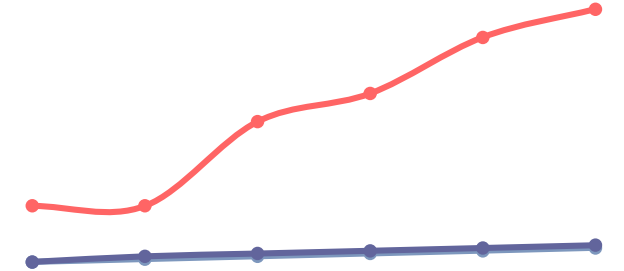


# AI, Bigdata & Cloud

ex) 코웨이

[http://www.dt.co.kr/contents.html?article\\_no=2018082402101631650002](http://www.dt.co.kr/contents.html?article_no=2018082402101631650002)

정수기와 공기청정기에 IoT를 적용, 수만대 기기에서  
분 단위로 데이터를 수집  
수집한 빅데이터는 이들 기기를 최적 상태로  
운영하고 디자인하는데 활용  
세계 최초로 머신러닝을 적용한 공기청정기를 출시,  
수집한 데이터를 클라우드 플랫폼에서 빠르게 분석해  
기기 운영에 바로 활용  
하이브리드 클라우드 모델



## 투자 현실 기대치

개발만 하고 싶은데  
행정이 참 많아요

요구하는 각종 문서들이  
주르르르륵

.....  
수퍼개발자들은 안들어오는 현실

취창업장려인데 이자놀이  
탁상행정

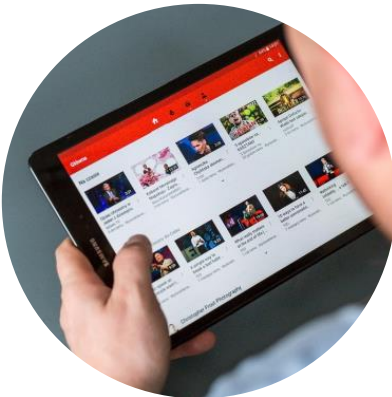
스타트업 장려는 하는데 국가 R&D  
사업에 스타트업 KPI가 전무

## 근원적인 문제점 대두

눈높이는 AWS, Google, MS Azure ... 인데  
투자는 그만큼 하는가?

비교대상의 문제 - -  
말하는 그들이 개발이 하던지  
그만큼 투자를 해도 될까 말까인데

PS) 우픈 현실



INNOCUS

# 멀티코어 -> 매니코어

## 매니코어 프로젝트

인프라가 안되는데 ....  
달나라가는 소리를 ....  
원천기술에 대한 고민



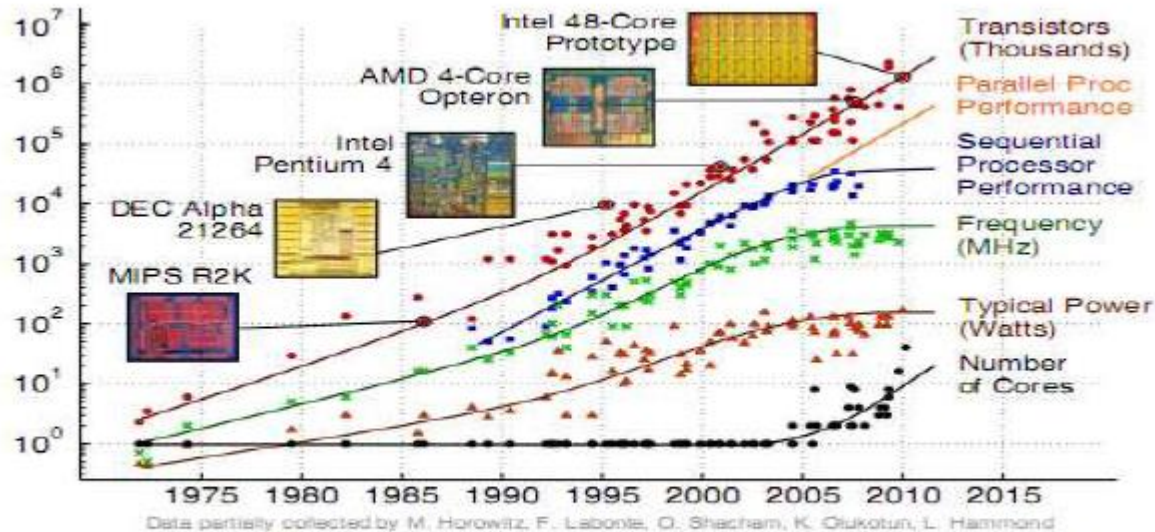
### 주파수의 한계

4G 5G CPU ?!



### 멀티코어의 한계

2.5G~3G Single, Dual, Quad.....Hexa....



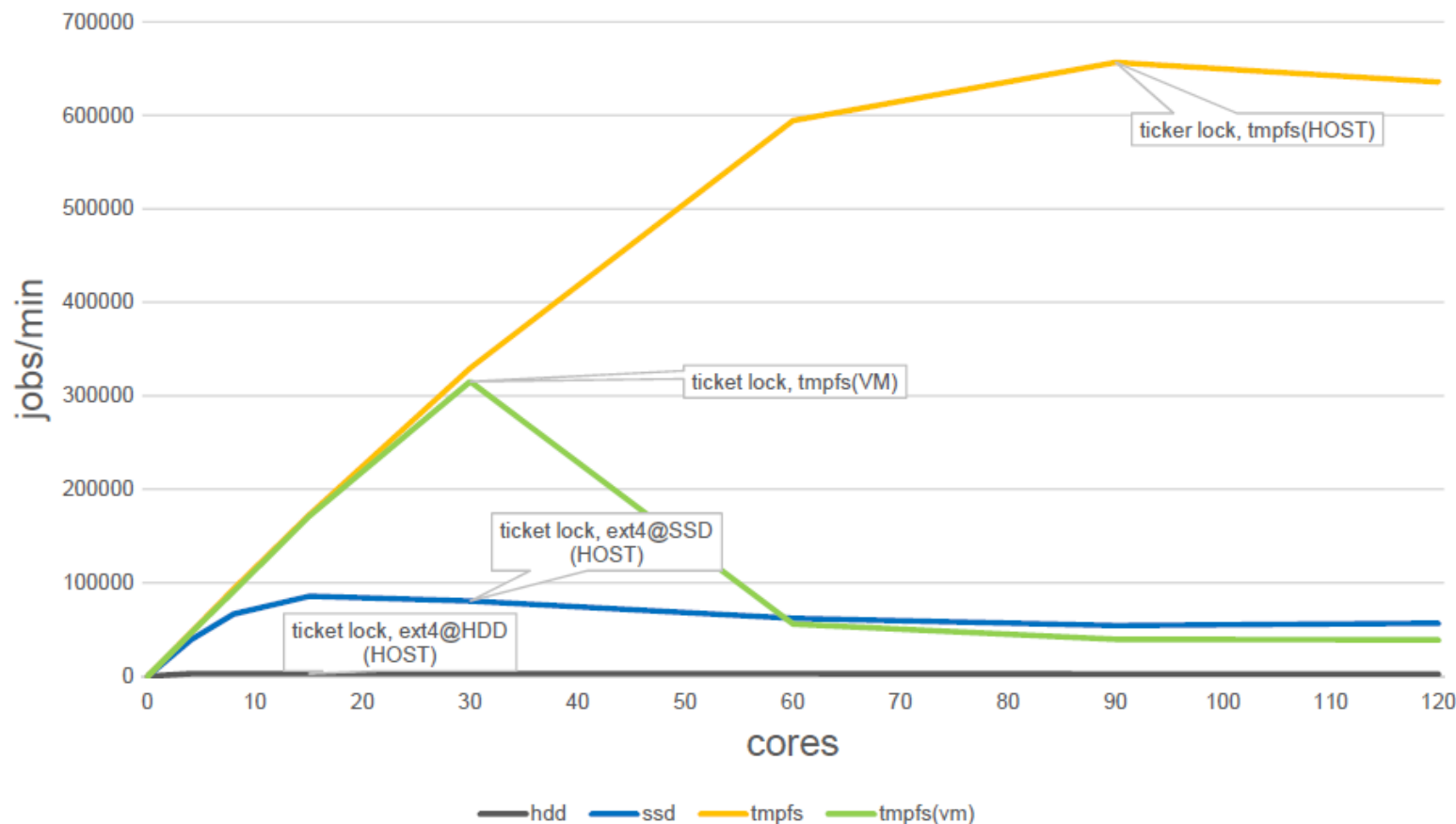
Prepared by C. Batten - School of Electrical and Computer Engineering - Cornell University - 2005 - retrieved Dec 12 2012 - <http://www.csl.cornell.edu/courses/ece5950/handouts/ece5950-overview.pdf>



INNOCUS

## 기존 SW 잘 동작하나? 기존 SW 성능이 잘 나올까 ?

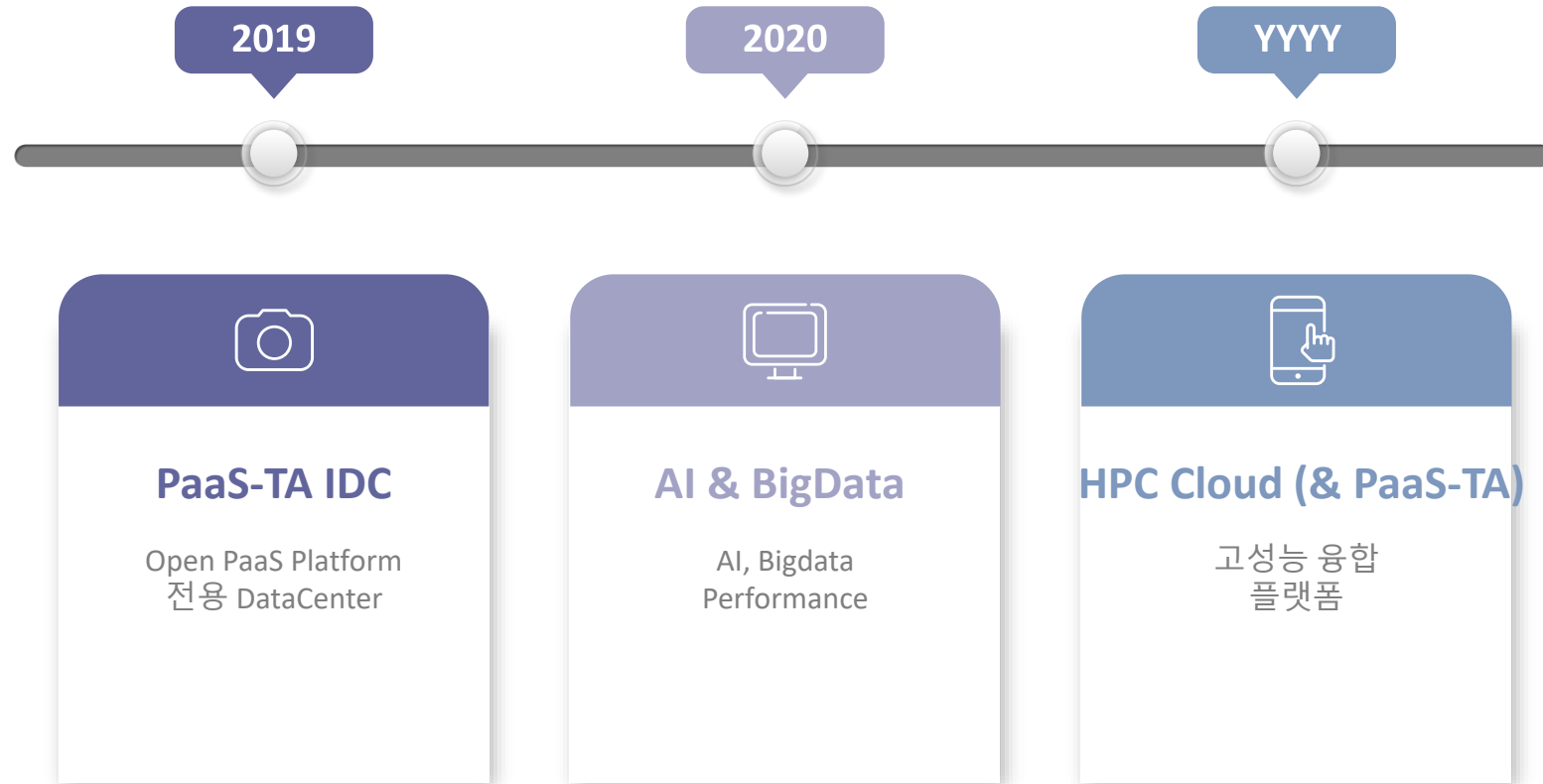
- AIM7(OS 성능시험 도구) Multiuser 워크로드(계산:IO=50:50)로 성능 실험



❖ 출처 : 리눅스커널 4.1버전@ETRI IBM 120코어 시스템(인텔)

# PROCESS DIAGRAM

PaaS-TA & AI, Bigdata on HPC Cloud





# THANK *you*



## LOCATION

321 114, seochojungang-ro, Seocho-gu,  
Seoul, Republic of Korea  
台北市 中山區 大直街 52巷  
2號 1樓  
原潭科技股份有限公司



## PHONE NUMBER

+10 7222 9981



## EMAIL

[ldbslovis@innocus.kr](mailto:ldbslovis@innocus.kr)



INNOCUS