

NAVER AI : Service & Platforms

Kwak Yong-Jae

Computer systems that can sense their environment, think, learn, and take action in response to what they're sensing and their objectives

- **Automated Intelligence:** Automation of manual/cognitive and routine/non-routine tasks
- **Assisted Intelligence:** Helping people to perform tasks faster and better
- **Augmented Intelligence:** Helping people to make better decision
- **Autonomous Intelligence:** Automating decision making processes without human intervention

Clova : AI Platform Sponsored by NAVER & LINE

**NAVER and LINE team up to develop a General-purpose AI platform
Synergies through Co-operation**



Clova inside Products

NAVER KOREA



Wave (2017)



Friends (2017)



Friends Minions (2018)



Friends Doraemon (2018)

LINE JAPAN



LINE
[スマートスピーカー] Clova Friends Brow



LINE
[スマートスピーカー] Clova Friends Sall

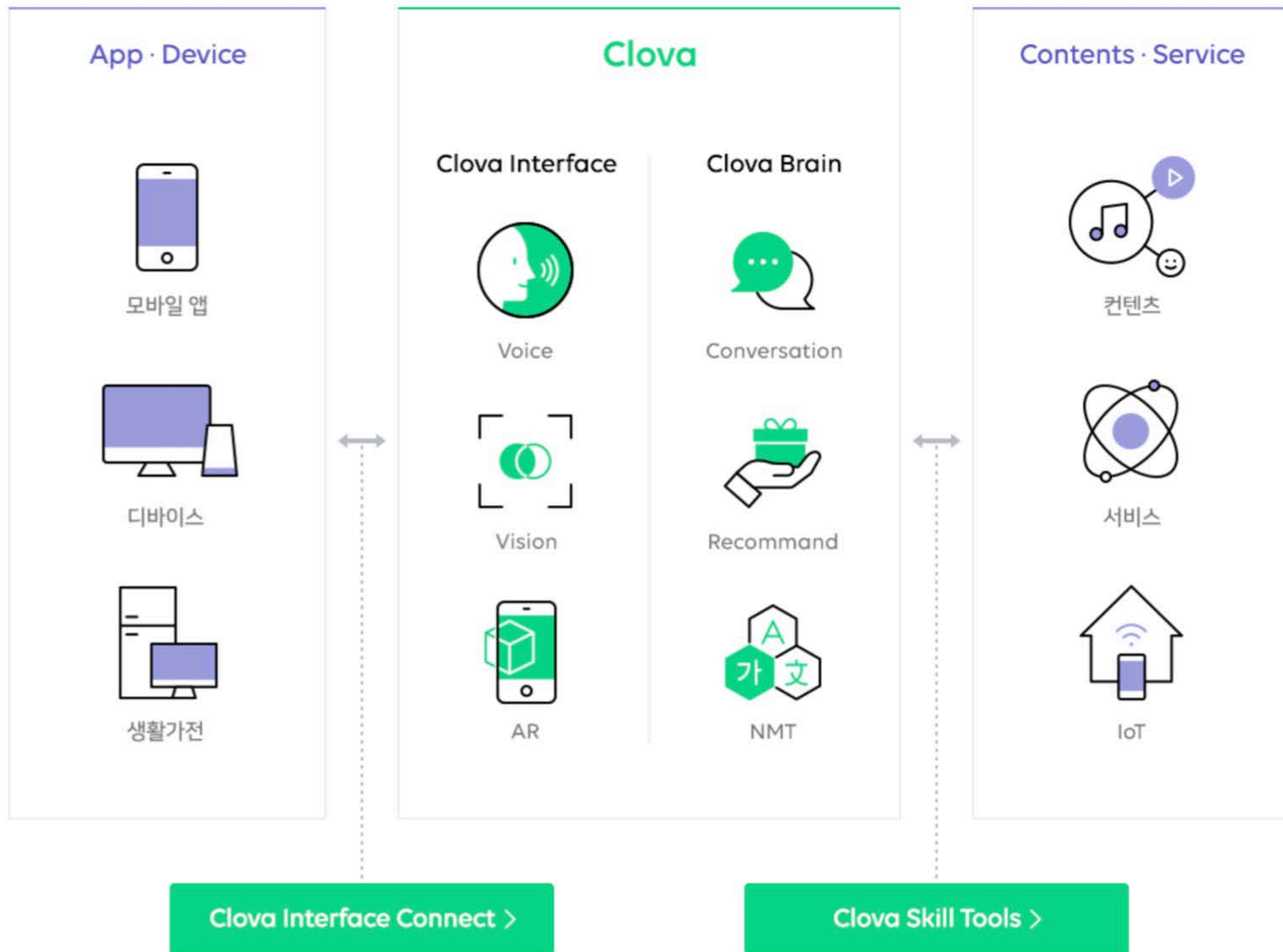


Friends Doraemon (2018)



Sony Earphone (2018)

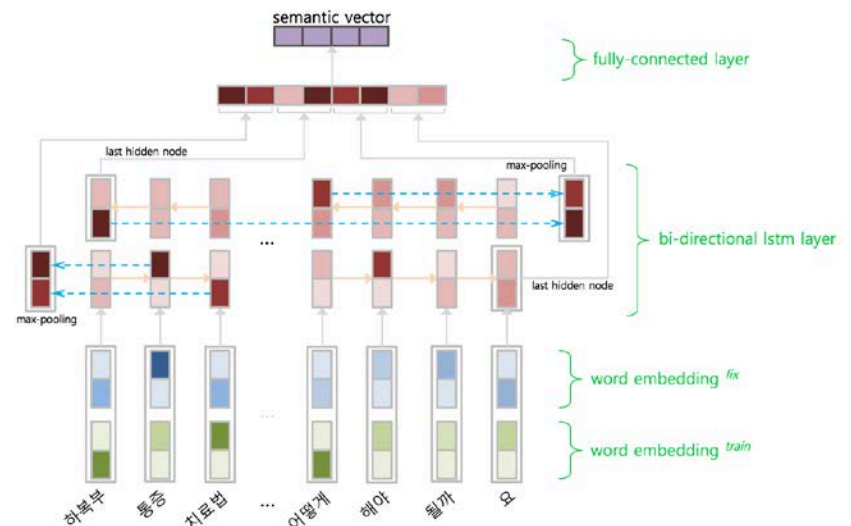
CLOVA Platform Service



AI Services: Enhanced NLP

Semantic based Matching

- Using 25M QA documents to embed a query to a vector
- Top score on Quora and SNLI dataset



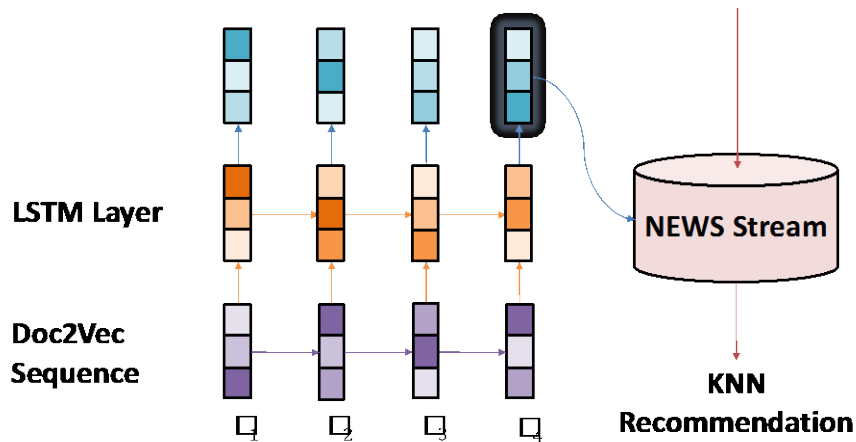
Kim, et. al, "Semantic Sentence Matching with Densely-connected Recurrent and Co-attentive Information", arXiv:1806.08730, 2018

Song, et. al, "Translation of Natural Language Query into Keyword Query Using a RNN Encode Decoder", SIGIR 2017

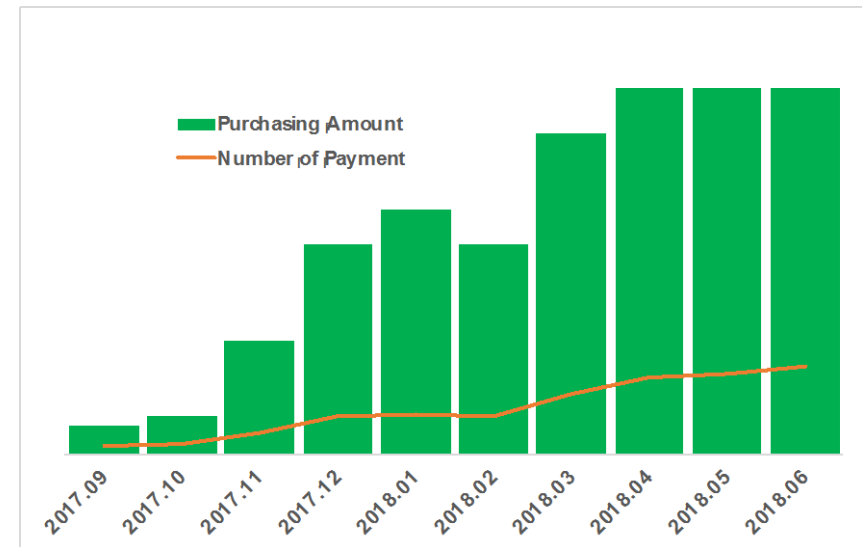
Kim, et. al, "Question Retrieval using Deep Semantic Matching for Community Question Answering", HCLT 2017

Leveraging Search Technology

- User Profiling and Contents Understanding
- AiRS: Collaboration Filtering, RNN-based News Recommendation
- AiTEMS: Personalized Shopping Recommendation



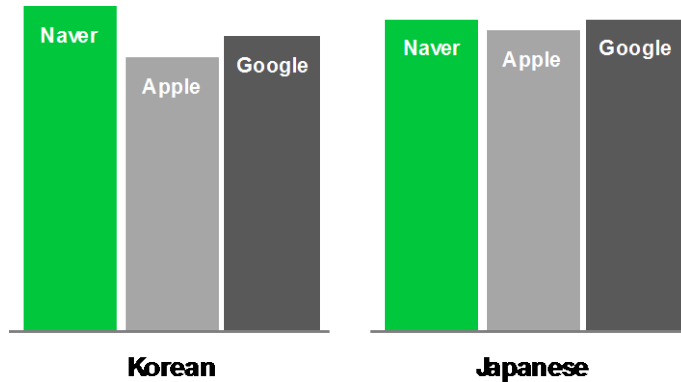
RNN-based News Recommendation



Purchasing Amount Increase through AiTEMS

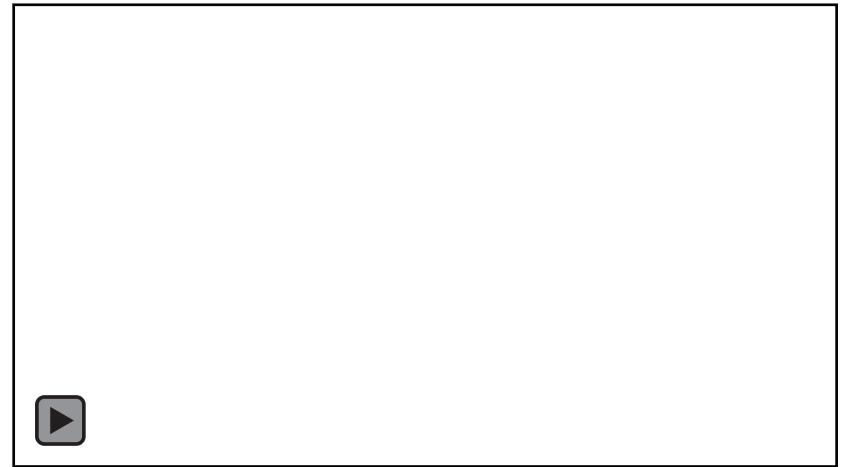
AI Services: VOICE & SPEECH

Up-to-date Dictionary and Language Model with 300million search queries per day

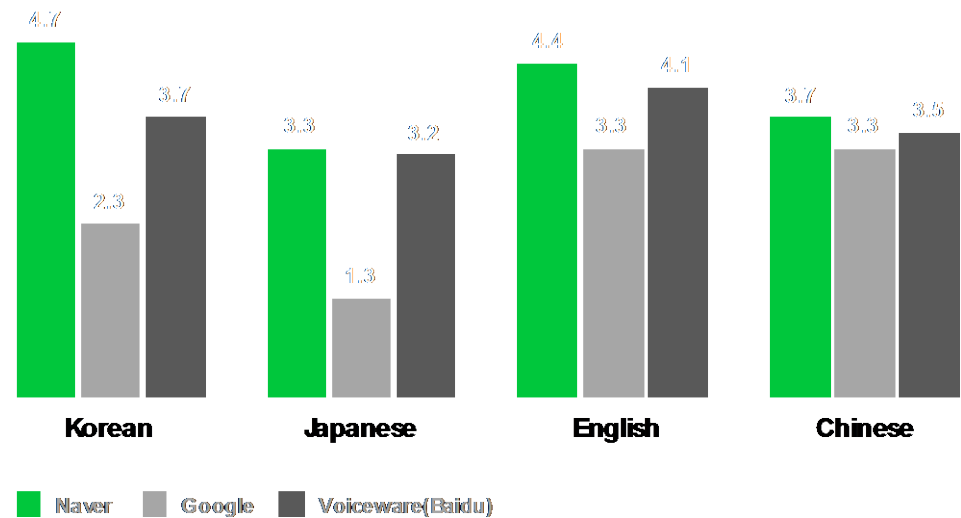
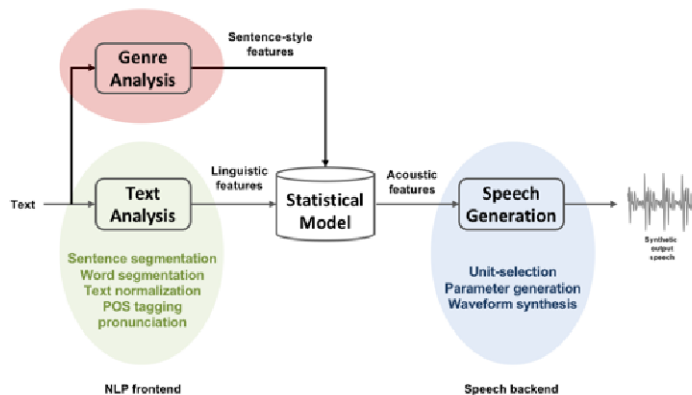


Accuracy of Speech Recognition

Data Set :
Papago's everyday conversations
Near smartphone, quiet environment



Many diverse things to read (News, Dictionary, Blog, Café, ...)



- 30 people (15 male, 15 female)
Naver, Google, Voiceware(Baidu)

Clova Voice

'Excellent'

HDTS

World-class Voice Synthesis Technology capable of 'best quality' with 4hrs of recording

'Easy'

NES

Personalized Speech Synthesis Technology that allows users to make 'easy' synthesizers with 8mins (100 sentences) recording

'Everybody'

VoiceMaker

Voice Synthesis System that allows anyone to use the Voice Synthesizer of their choice

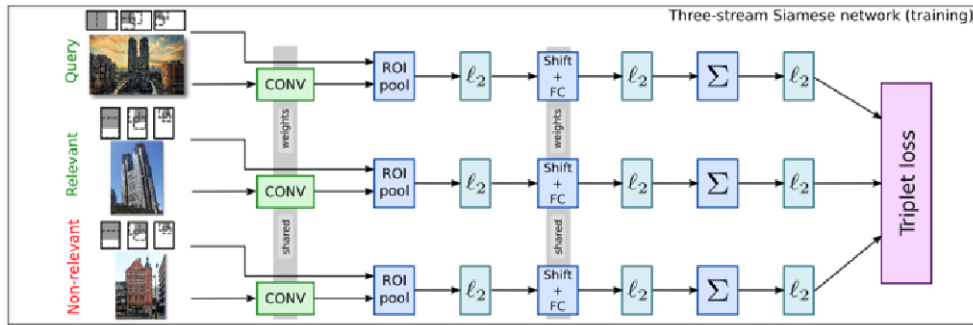
CLOVA Voice (HDTS Demo)

Generate personalized voice with 4-hour training data
- 1/10 training data reduction

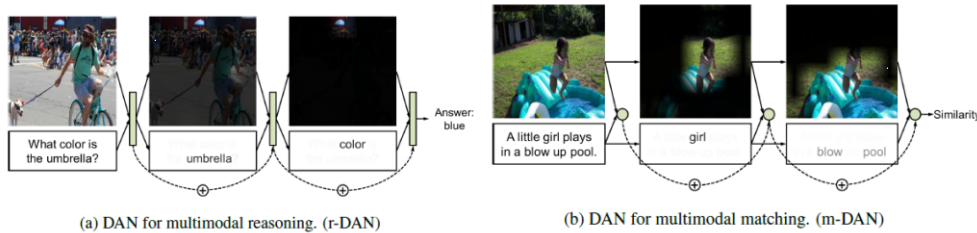


AI Services: VISION

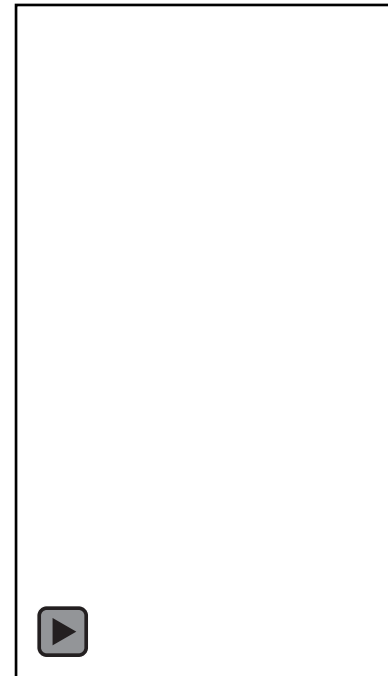
Smart Lens: DNN-based Visual Search Collaborating with Naver Labs Europe (Fashion, Visual, Search, and so on)



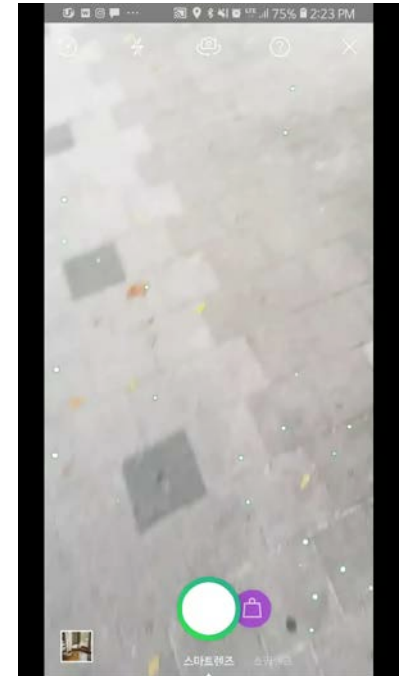
DNN-based Image Recognition Framework



Clova & LABS Europe



Shopping lens

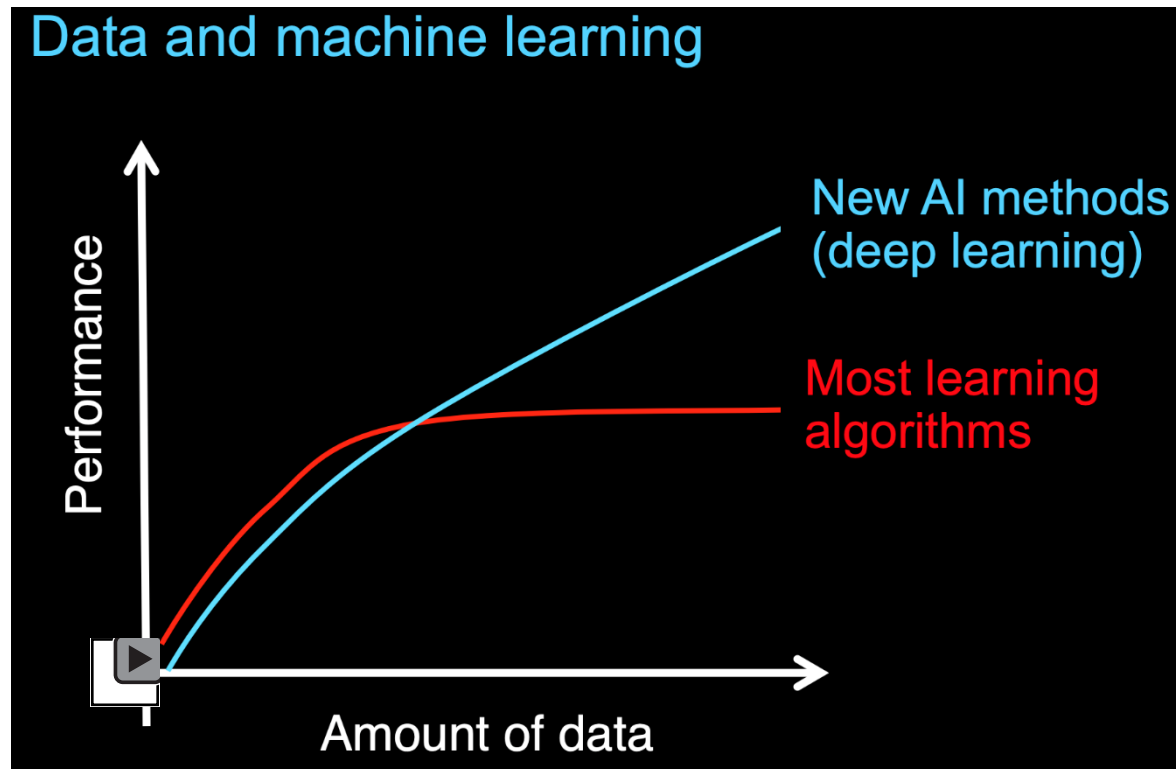


Smart lens

AI Service Platforms

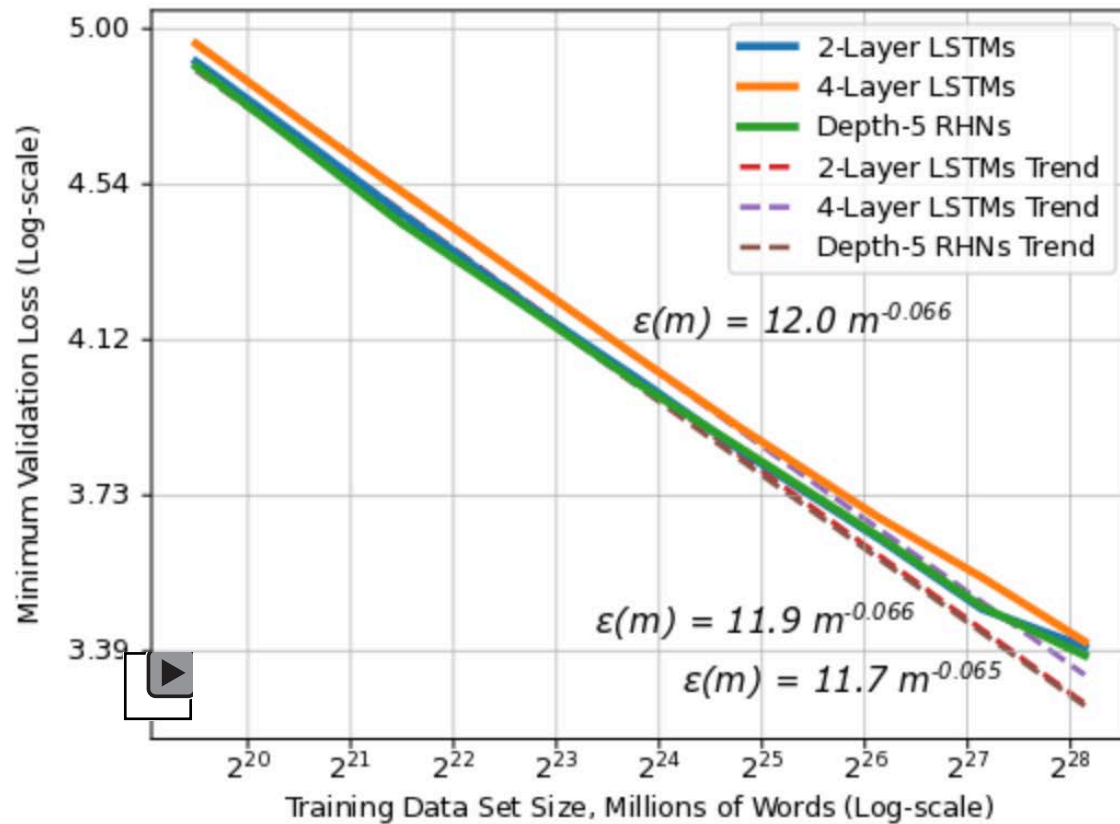
MORE DATA for a few Accuracy Improvement

(from Andrew Ng's observation)



MORE DATA for a few Accuracy Improvement

(from BAIDU Research)



“Enterprises must unify their data warehouse”

Andrew Ng,
Nuts & Bolts of Applying Deep Learning



UNIFYING DATA : FOUR essential features to be integrated for AI & ML

BACKUP & DATA WAREHOUSE

- key requirement: High Throughput File & Object

DATA LAKE serves

- Huge repository of raw data (contents and logs)
- Catalog for data retrieval
- Universal data Input & output API

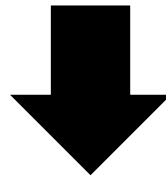
COMPUTING CLUSTER provides

- Computing resource based on distributed huge file system
- Big-data processing tools and application management env.

AI Service Engine Provides





- Deep-learning work-bench
- Inference PaaS for high-performance service production

기계학습과 인공지능 분야의
많은 문제가 Engineering에서 발생

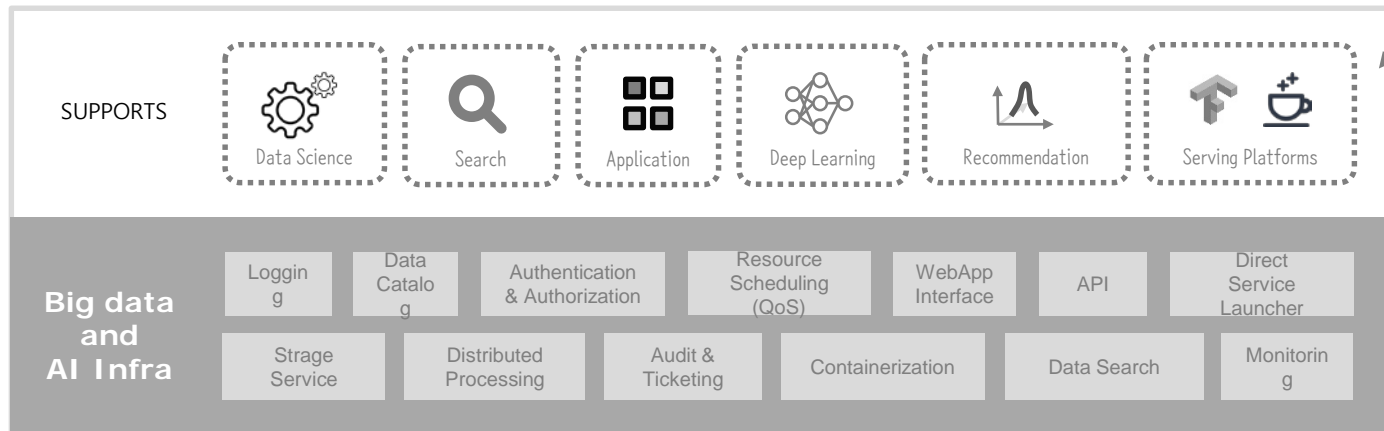


조율된 데이터 파이프라인
효율적인 분산처리 인프라

Big DATA and AI Infra @NAVER

 **Data Scientist**  **DL Researcher**  **Developer**  **Researcher**

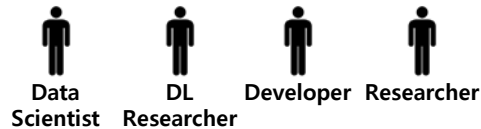
/* 이런 것들을 할 수 있도록 */



/* 이런 것들을 만들어 지원합니다.
그래서 이게 어떤 모습이나면, */

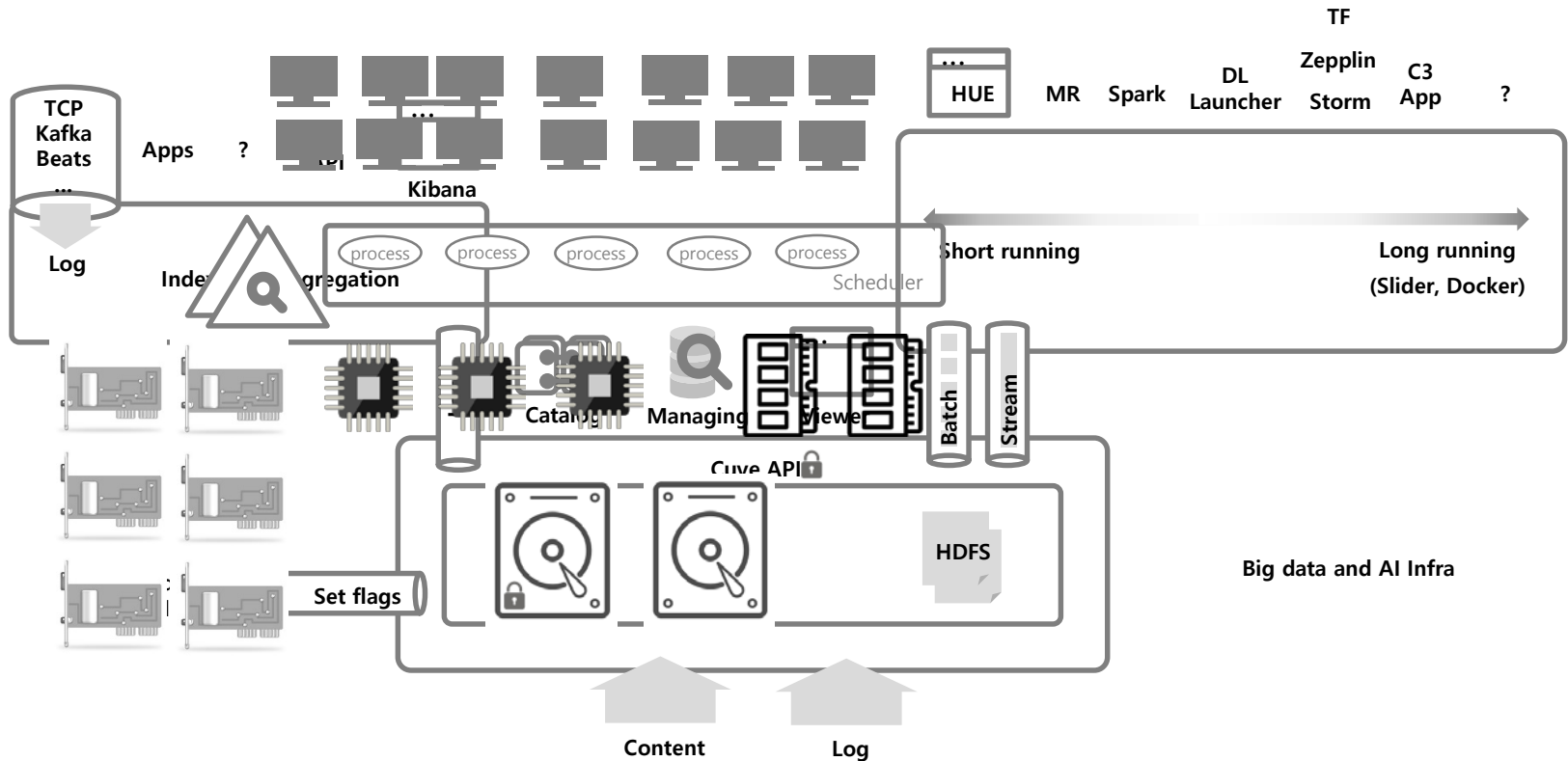
Big DATA and AI Infra @NAVER

/* 큰 컴퓨터라 보시면 됩니다.
모니터여러개, 운영체제(스케줄러),
CPU, 메모리, 하드, 네트워크카드에
앱도 돌리는.. */

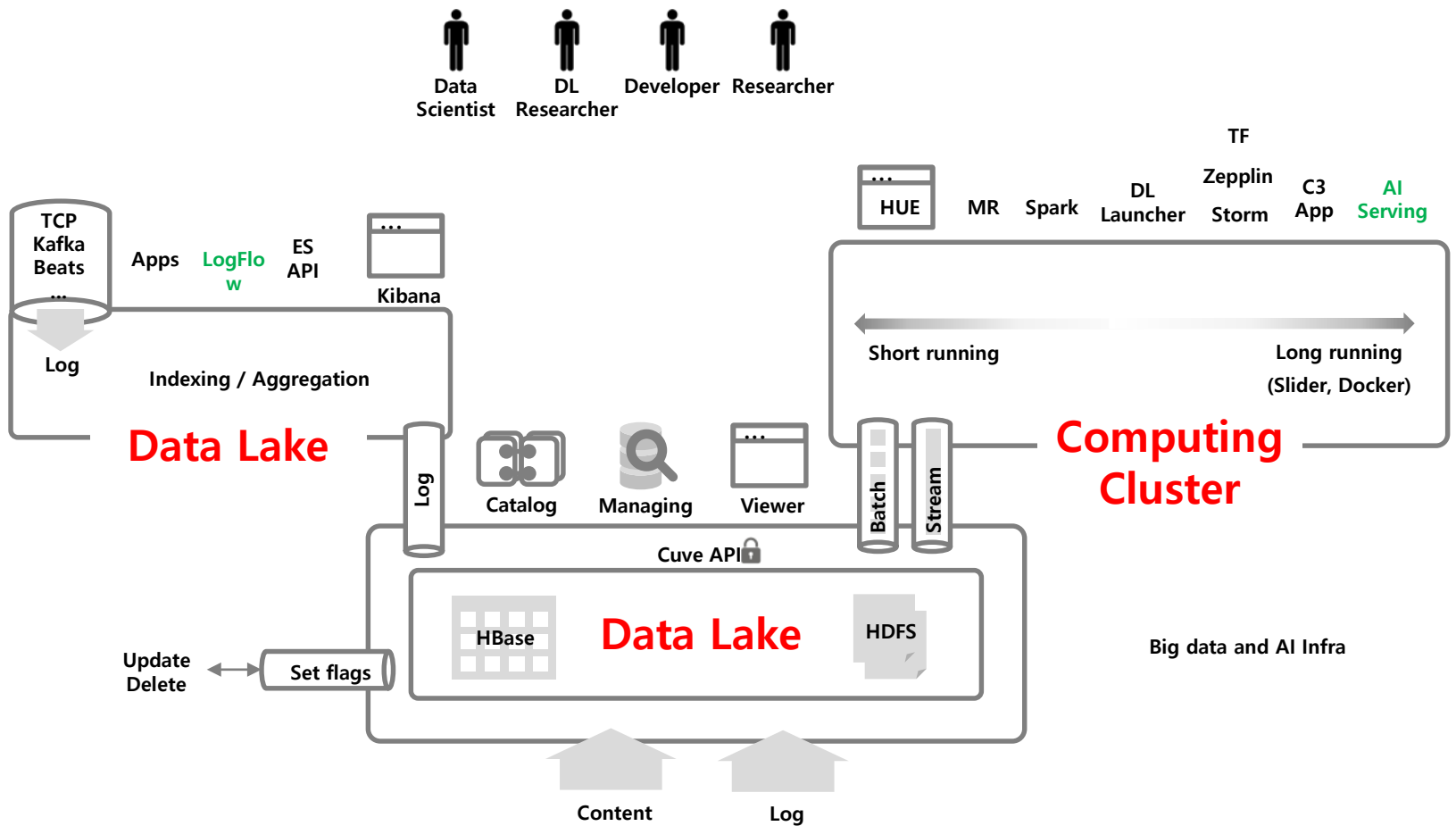


/* 요걸 살짝 움직이면, */

/* 이렇게 됩니다. 요게 data lake,
Computing cluster입니다 */



Big DATA and AI Infra @NAVER



UNIFYING DATA : Big-Data & AI Infra @ NAVER

BACKUP & DATA WAREHOUSE

- Provided by NAVER Infra-structure Group (Data-center Level)
- So, We do not care of it

DATA LAKE

- **CUVE**: Huge data cluster constructed based on hBASE
- **LOGISS**: Short-term Log-Ingester based on Elastic-Search

COMPUTING CLUSTER

- **C3**: Multi-tenant Computing Cluster based on Hadoop whose SIZE is Asia Big 3
- C3App: Integrated work-bench provided by C3

AI Service Engine

- **C3-DL**: Deep-learning toolkit based on Nvidia-docker & GPU, CPU
- **AISP**: Inference PaaS based on Tensor-Flow Serving (planned PyTorch)

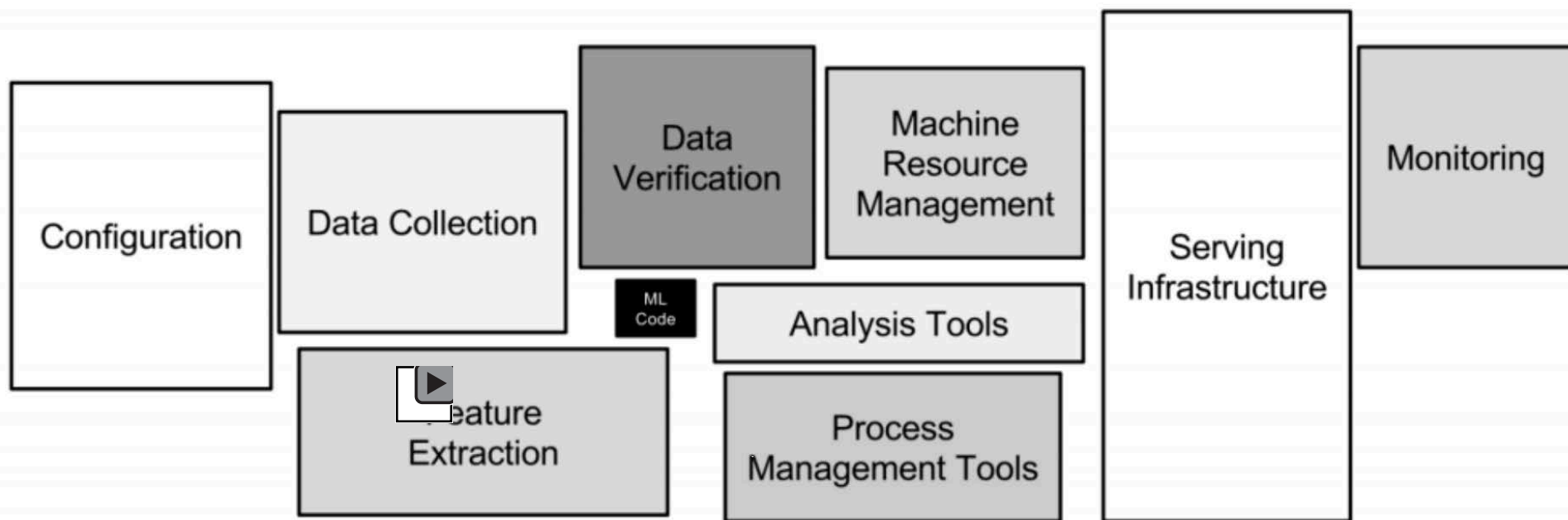
Big DATA and AI Infra @NAVER

플랫폼		분류	데이터	사용량
Cuve	컨텐츠	Data Lake	X천억 건 Y PB	초당 1.5 ~ 2만 문서 유입 초당 70만 ~ 200만 문서 배출
	로그		X조 건 ZZ PB	초당 90억 건 유입 초당 1,000억 건 배출
Logiss		Data Lake	X백억 건 ZZ TB	초당 20만 개 로그 색인 900개 색인 운영
C3		Computing Cluster	X억 개 ZZ PB	동시 실행 작업 : 1500 개 하루 평균 실행 작업 : 20만 개
C3 DL		AI Service		동시 실행 작업 : 120 개 하루 평균 실행 작업 : 1천 개
AiSP		AI Service		하루 3~4억 개 인퍼런스 수행 중

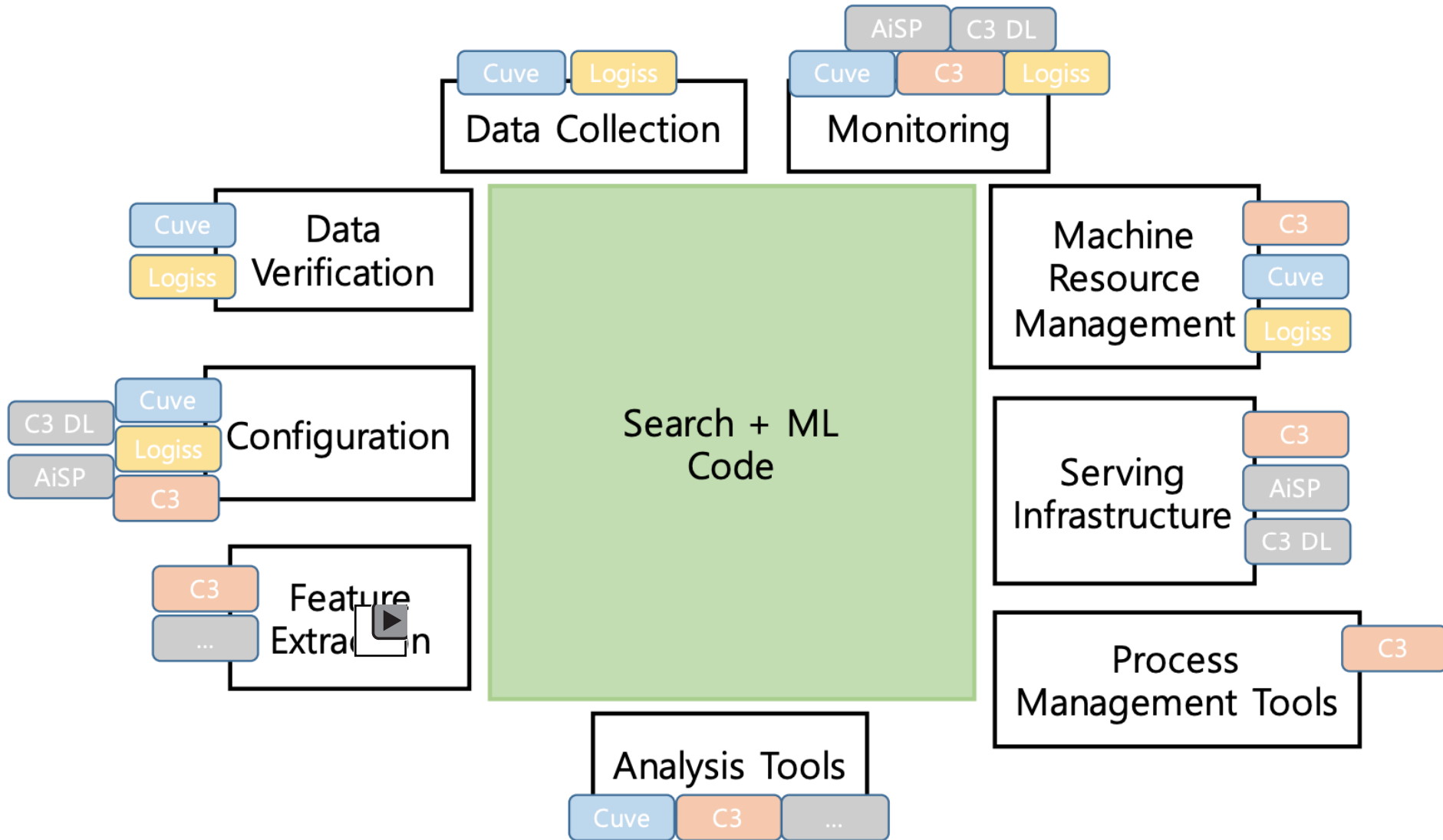
AI & ML Owe so many technical depts to ENGINEERING!!

Hidden Technical Debt in Machine Learning Systems

D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips
`{dsculley, gholt, dgg, edavydov, toddphillips}@google.com`
Google, Inc.



Correspondance b/w Technical Debts and NAVER Big-Data & AI Infra



Example:

AI @ Shopping Search

AS-WAS: 롱테일 질의에 대한 쇼핑 검색 결과가 좋지 않음

질의에 분류가 매칭될 확률을 구하는 모델이 이미 있으나,
롱테일에서는 클릭수 및 질의수 부족으로 확률이 0인 경우가 많음

잘못된 분류 상품

화질 좋은 티비

가격 + 혜택 +
색상 + 상품타입 +
판매처 + 옵션 더보기

네이버 랭킹순



LG미니빔TV PF50KA 600인치...
799,000원
무료배송 | 리뷰 9 | 구매 13
클레이윈도
체크포인트
포인트 7,990원

#생동감넘치는 #선명한화질 #슬림한디자인



LG미니빔TV PH50HT /PH550
블랙 버전 550인치 베스트 셀...
519,000원
무료배송 | 리뷰 23 | 구매 47
체크포인트
포인트 5,190원

#휴대하기좋은크기 #슬림한디자인 #선명한화질




LG미니빔TV PH50HT /PH550...
519,000원


가볍고 저렴한 자전거

가격 + 혜택 +
색상 + 상품타입 +
판매처 + 옵션 더보기


네이버 랭킹순



(S11)가볍고 시원한 레이온 애플 자전거 홈웨어 5부바
9,050원
G마켓



가볍고 통기성좋은 메쉬등판 다용도 허리백 어깨스트랩 탈...
최저 41,600원
판매처 2



CREE XM-L2 USB라이트 올블랙 랜턴
13,800원
리뷰 365 | 구매 887
EMALL

검색결과 0건

에르스메스 스카프

검색결과가 없습니다.
다른 검색어를 입력해 보거나,
옵션을 재선택해 보시기 바랍니다.

일시적으로 상품이 품절되거나 노출이 제한된 상품일 수 있습니다.

검색어의 질자가 정확한지 확인해 보시기 바랍니다.

옵션 항목을 다시 선택해 보시기 바랍니다.

이전 페이지로 가기

누룽지 굽는 기계

검색결과가 없습니다.
다른 검색어를 입력해 보거나,
옵션을 재선택해 보시기 바랍니다.

일시적으로 상품이 품절되거나 노출이 제한된 상품일 수 있습니다.

검색어의 질자가 정확한지 확인해 보시기 바랍니다.

옵션 항목을 다시 선택해 보시기 바랍니다.

이전 페이지로 가기

법적고지

자세히 보기 >

네이버(주)는 결제정보의 중개서비스 또는 통신판매중개시스템의 제공자로서, 통신판매의 당사자가 아니며 상품의 주문, 배송 및 환불 등과 관련한 의무와 책임은 각 판매자에게 있습니다.

현금 결제 시 주의사항

자세히 보기 >

법적고지

자세히 보기 >

네이버(주)는 결제정보의 중개서비스 또는 통신판매중개시스템의 제공자로서, 통신판매의 당사자가 아니며 상품의 주문, 배송 및 환불 등과 관련한 의무와 책임은 각 판매자에게 있습니다.

현금 결제 시 주의사항

자세히 보기 >

TO-BE: 롱테일 질의에 대해서도 분류매칭 확률을 구할수 있게 하자

[질의-속성] 관계 추출 모델을 사용해 **실시간으로** 질의-카테고리 매칭 확률 계산
모든 질의에 대해 질의매칭 확률 계산 가능

화질 좋은 티비

가격	+	혜택	+
색상	+	상품타입	+
판매처	+	옵션 더보기	▼

네이버 랭킹순 ▼

NCMP 테스트 결과

카테고리: TV 0.987

카테고리 내 검색 결과

전체 > 디지털/가전 > 영상가전 > TV

LEDTV

접기 ^

네이버 랭킹순 ▼

LG미니빔TV PF50KA 600만...
799,000원
무료배송 | 리뷰 9 | 구매 13
플레이윈도
체크포인트
포인트 7,990원

#생동감넘치는 #선명한화질 #슬림한디자인

LG미니빔TV PH50HT /PH550
블랙 버전 550만시 베스트 셀...
519,000원
무료배송 | 리뷰 23 | 구매 47
체크포인트
포인트 5,190원

#휴대하기좋은크기 #슬림한디자인 #선명한화질

LG미니빔TV PH50HT /PH550...
519,000원

[더함] 55인치 UHD TV 코스모 ...
419,000원
리뷰 38 | 구매 66
플레이윈도
더함 TV 공식 온라인몰
포인트 4,190원

가볍고 저렴한 자전거

가격	+	혜택	+
색상	+	상품타입	+
판매처	+	옵션 더보기	▼

네이버 랭킹순 ▼

NCMP 테스트 결과

카테고리: 자전거/MTB 0.989

카테고리 내 검색 결과

전체 > 스포츠/레저 > 자전거 > 자전거/MTB

사이클 하이브리드
미니벨로 전기자전거
일반자전거 MTB

접기 ^

자전거/MTB ×

네이버 랭킹순 ▼

(511)가볍고 시원한 레이온 애
플자전거 홈웨어 5부바
9,050원
G마켓

가볍고 통기성좋은 메쉬등판
다용도 허리색 어깨스트랩 탈...
최저 41,600원
판매처 2

CREE XM-L2 USB라이트 올블
랙 랜턴
13,800원
리뷰 365 | 구매 887
EM몰
포인트 4,300원

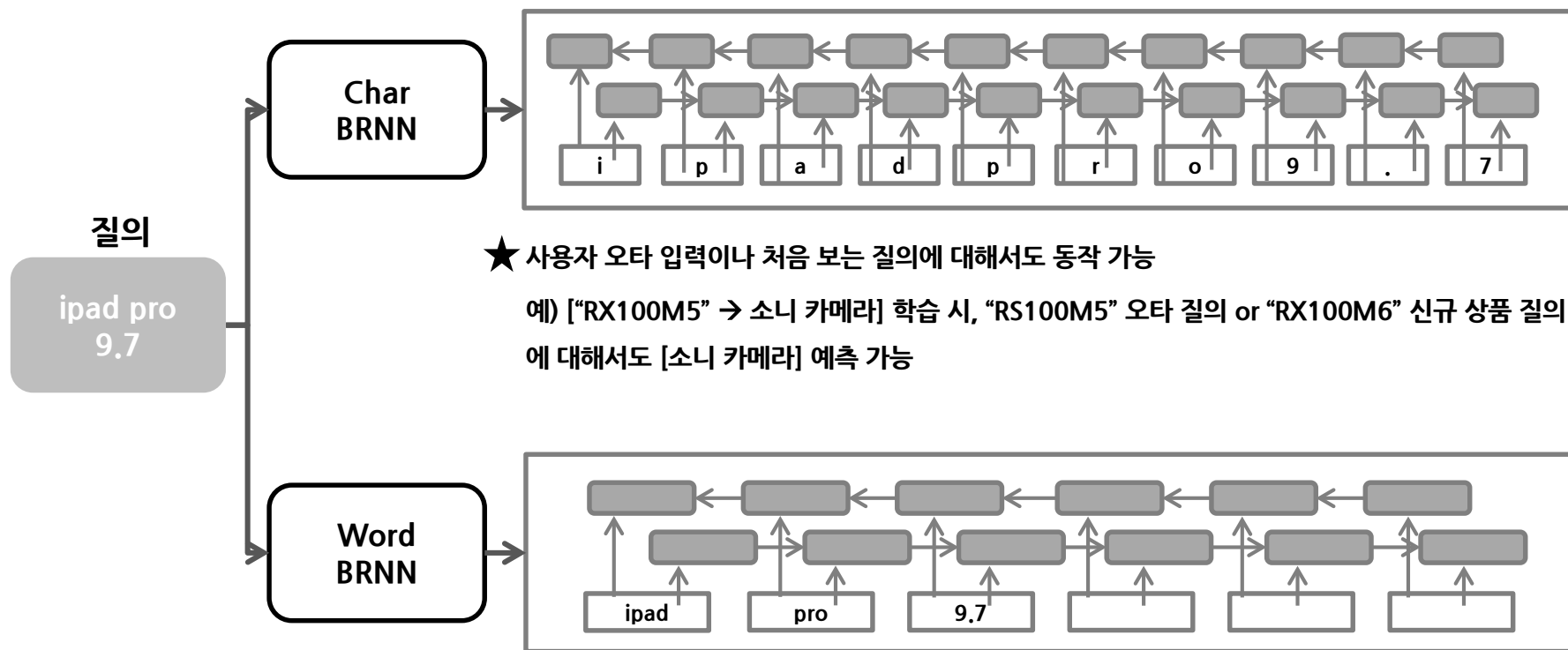
이월상품 할인 2017 메리다 스
컬트라 300 티아그라 입문형...
920,000원
무료배송
스포츠존
포인트 9,200원

#가볍고세련된 #하체를잡아주는사이클

Approach: Machine Learning using Neural Network

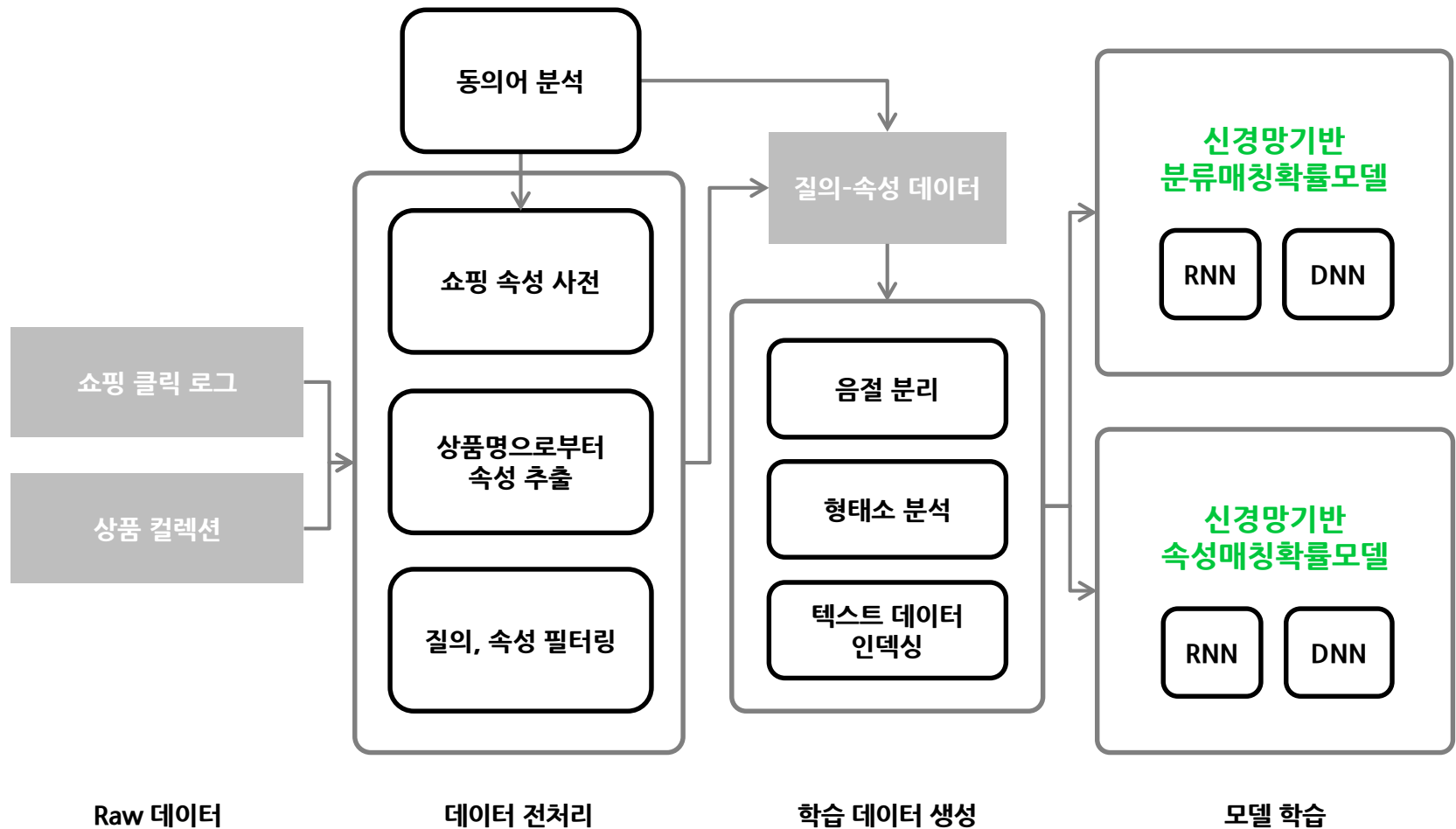
Long Short Term Memory - Bidirectional Recurrent Neural Networks

음절 단위 / 단어 단위 동시 학습



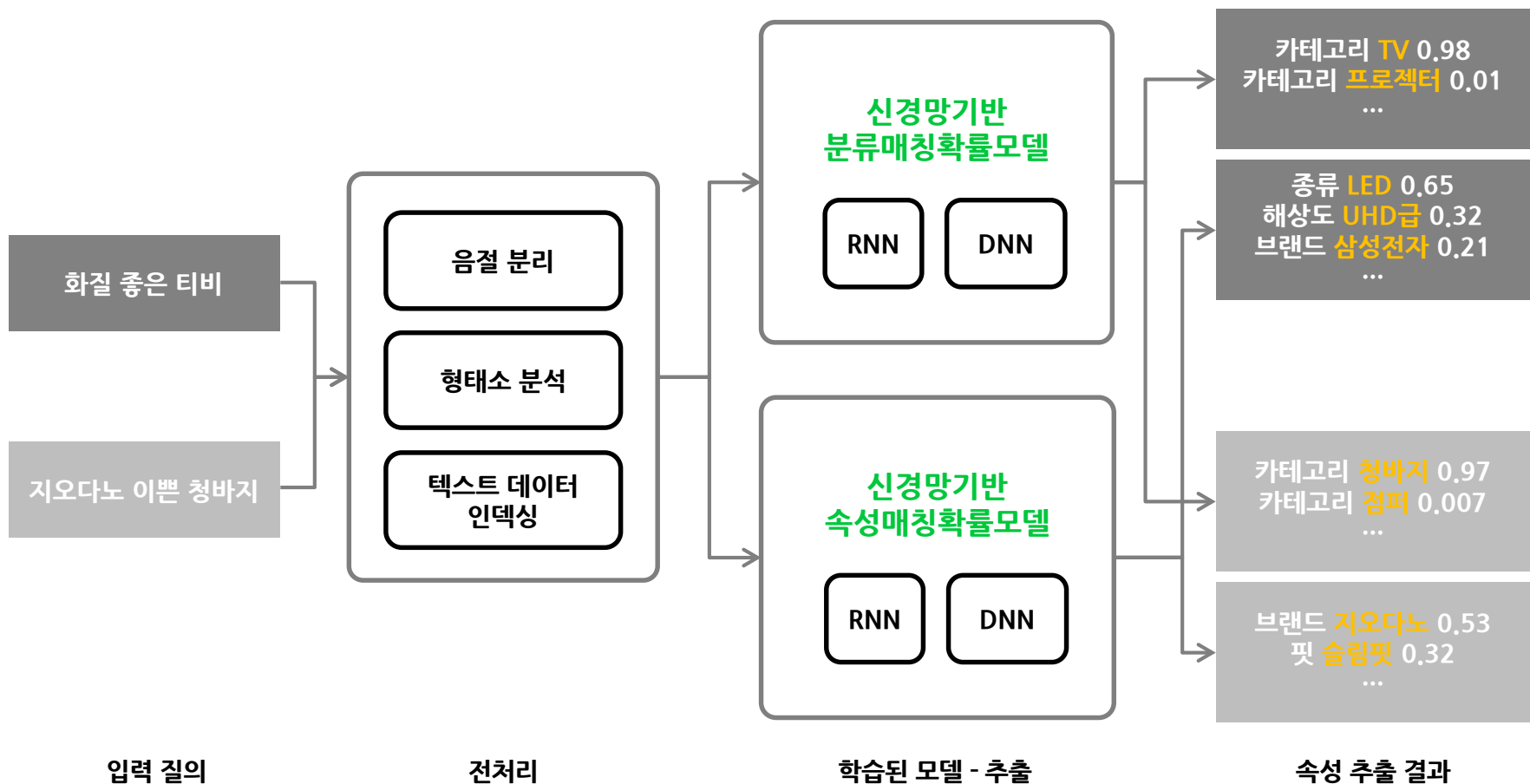
Approach: Machine Learning using Neural Network

Model Learning Process



Approach: Machine Learning using Neural Network

Inference Process using Learned Model



Implement: Processing at Big-Data & AI Infra Side

질의어에 적합한 쇼핑 카테고리 인퍼런스

← → ↻ ⓘ csh60329.nfra.io:51337/ncmp.aps?q=원피스

This XML file does not appear to have any style information as

DEMO:

<http://stage.ncmp-aps.kr-cc.fe.iqe.navercorp.com/?q=%EC%9B%90%ED%94%BC%EC%8A%A4>

← → ↻ ⓘ csh60329.nfra.io:51347/nexp.aps?q=문재인

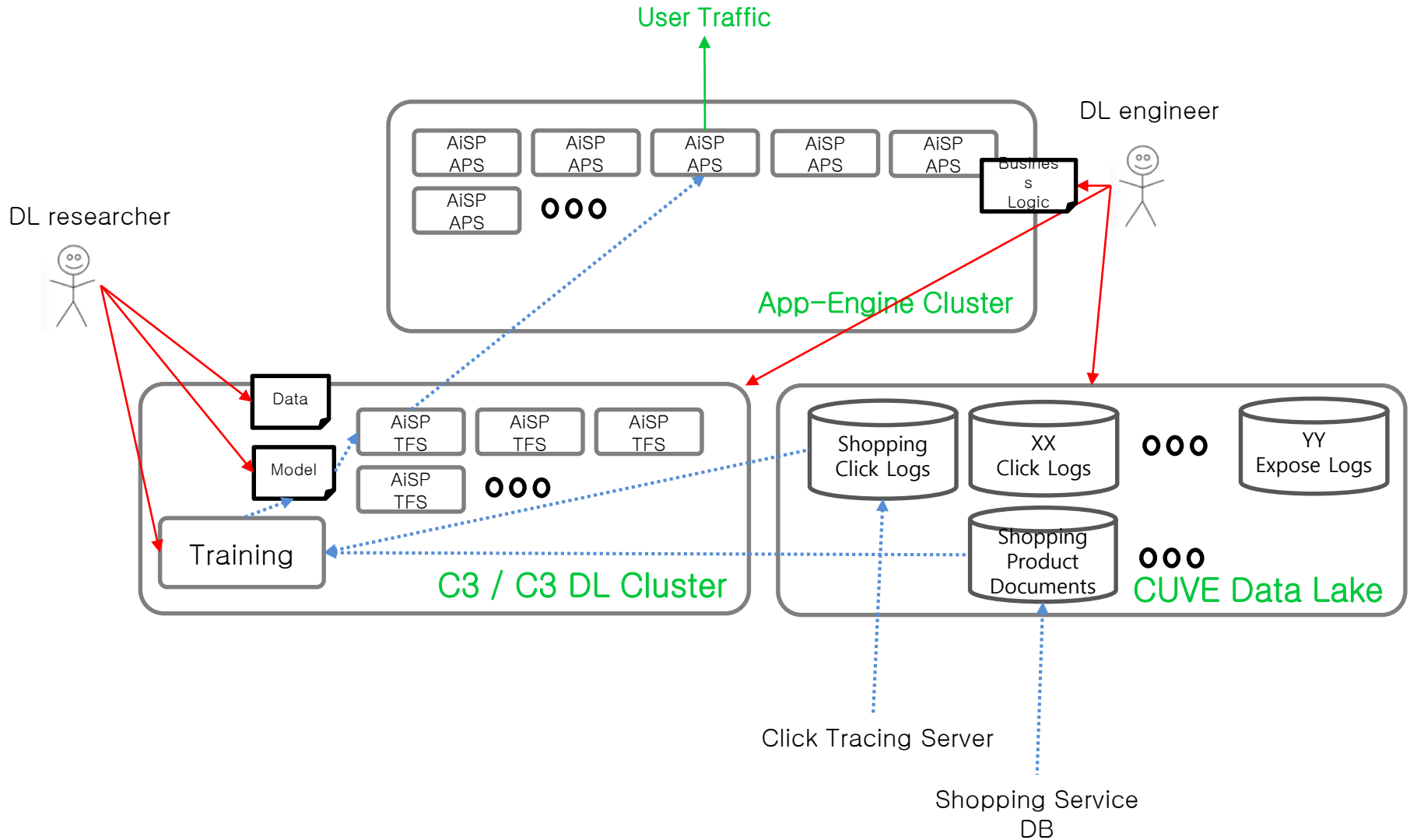
This XML file does not appear to have any style information associated with

```
<?xml version="1.0" encoding="UTF-8"?>
<result>
  <prob_service>
    <item rank="1" probability="0.341922" label="9">nws</item>
    <item rank="2" probability="0.305854" label="1">fsn</item>
    <item rank="3" probability="0.165765" label="12">vdo</item>
    <item rank="4" probability="0.120078" label="2">img</item>
    <item rank="5" probability="0.05236" label="3">kdc</item>
    <item rank="6" probability="0.0044636" label="11">sit</item>
    <item rank="7" probability="0.00369693" label="0">bok</item>
    <item rank="8" probability="0.00320936" label="7">mgz</item>
    <item rank="9" probability="0.001169" label="8">mus</item>
    <item rank="10" probability="0.000724829" label="5">ldc</item>
  </prob_service>
  <status>
    <code>OK</code>
    <message/>
  </status>
</result>
```

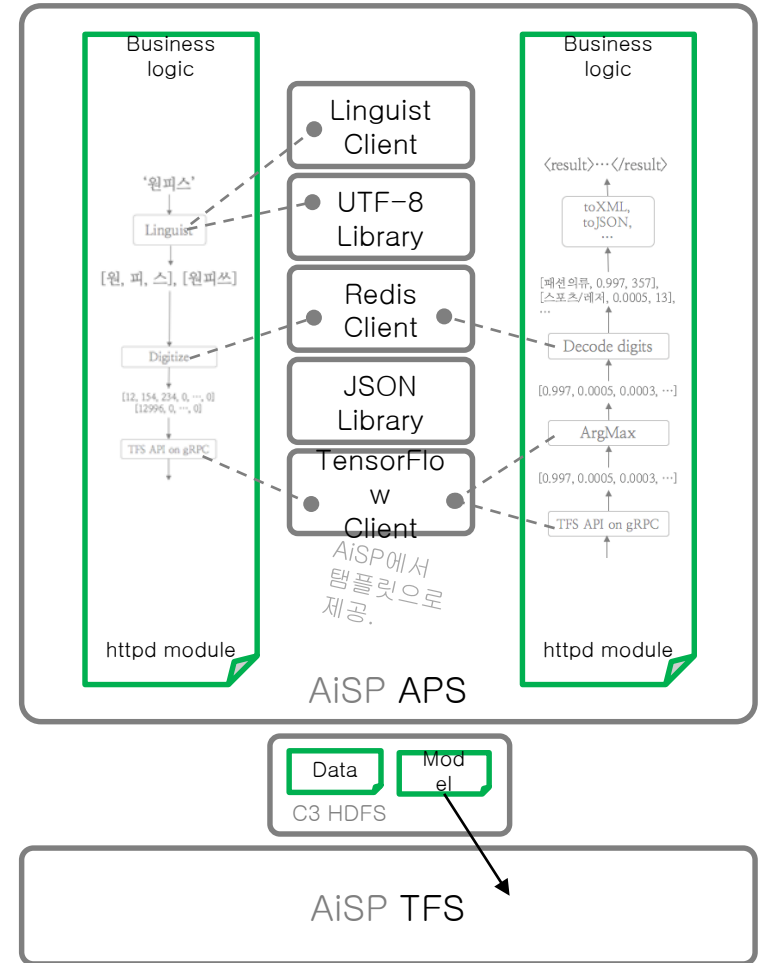
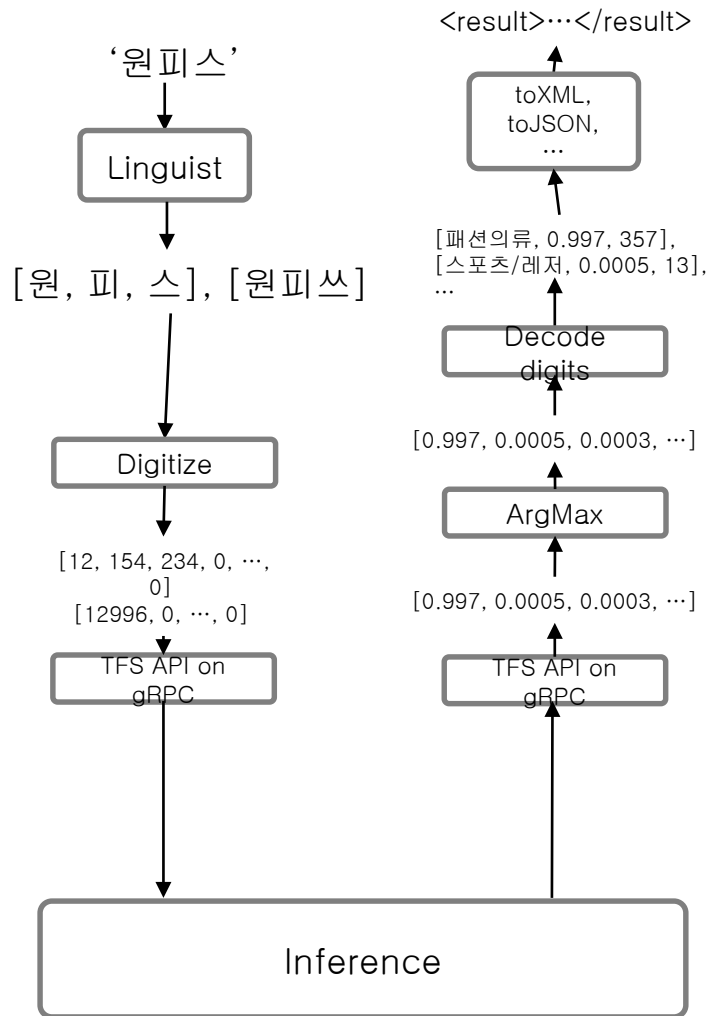
질의어로 특정 카테고리 노출 여부 판단

```
<?xml version="1.0" encoding="UTF-8"?>
<result>
  <prob_category1>
    <item>
      <rank>1</rank>
      <probability>0.997623</probability>
      <label>8</label>
      <group>category1</group>
      <name>패션의류</name>
      <code>50000000</code>
    </item>
    <item>
      <rank>2</rank>
      <probability>0.000932</probability>
      <label>4</label>
      <group>category1</group>
      <name>스포츠/레저</name>
      <code>50000007</code>
    </item>
    <item>
      <rank>3</rank>
      <probability>0.000513</probability>
      <label>3</label>
      <group>category1</group>
      <name>생활/거각</name>
    </item>
  </prob_category1>
</result>
```

Implement: Processing at Big-Data & AI Infra Side



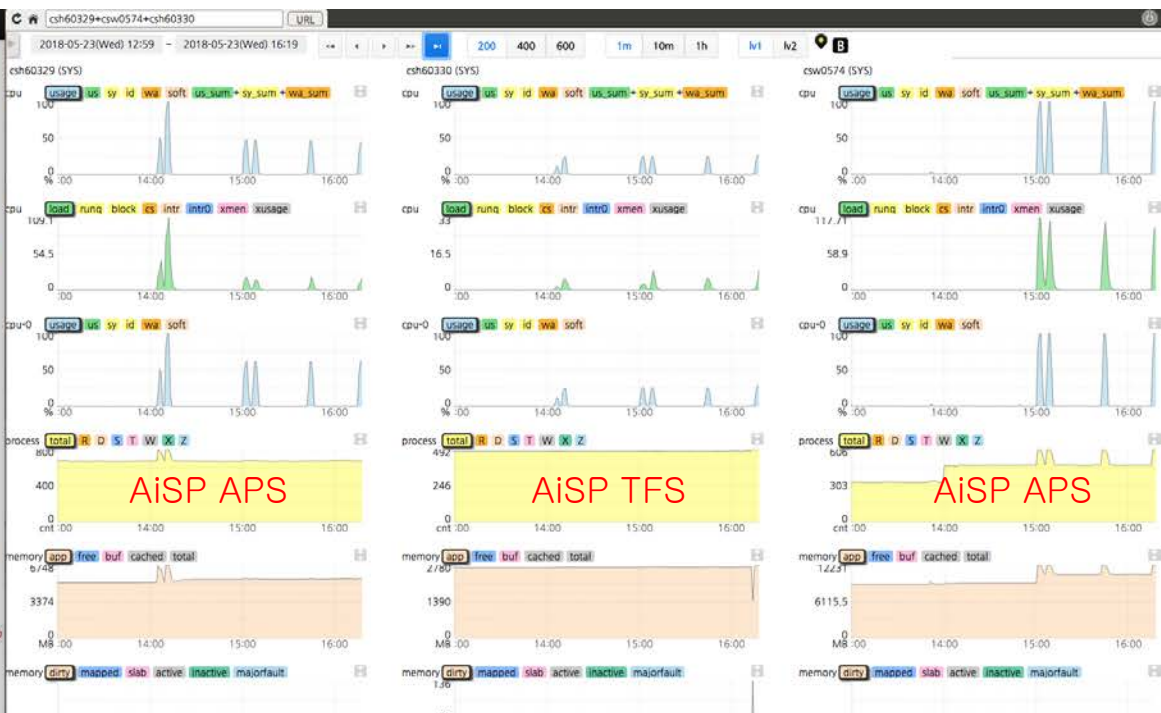
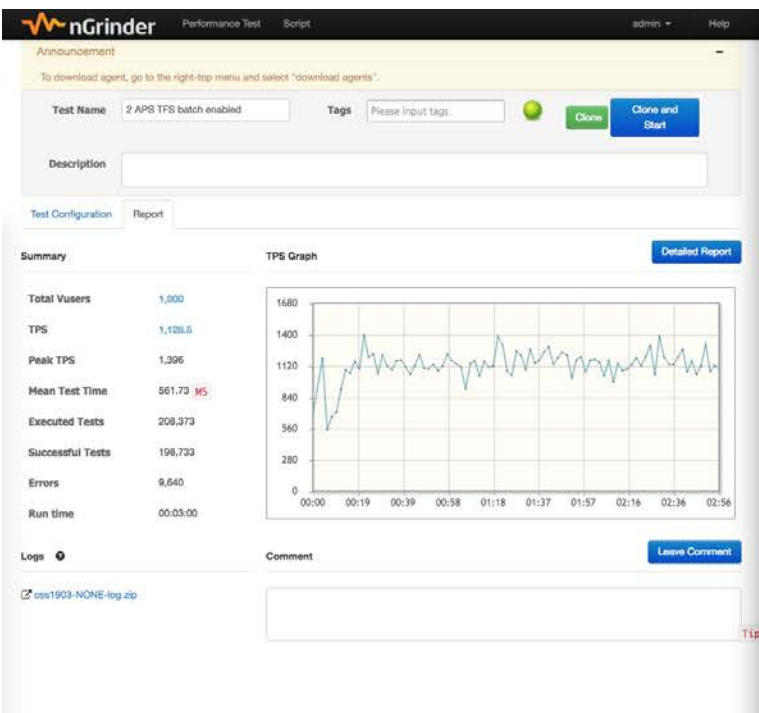
Implement : Processing at Big-Data & AI Infra Side



Implement : Processing at Big-Data & AI Infra Side

CPU 서버 48대로 가능.
심지어 튜닝하면 더 줄어든 가능성 있음.

Latency: TFS 130ms, APS 10ms



2.4GHz 16 cores

2.4GHz 16 cores

2.2GHz 12 cores

-

End of Document

-

Thank You.