

# Connected Data Architecture 개념의 확장을 통한 AI 서비스 초안 설계

(Draft Design of AI Services through Concept Extension of Connected Data Architecture)

차병래<sup>\*,\*\*</sup>, 박선<sup>\*</sup>, 오수열<sup>\*\*\*</sup>, 김종원<sup>\*</sup>

(ByungRae Cha, Sun Park, Su-Yeol Oh and JongWon Kim)

## 요약

DataLake 프레임워크와 같은 단일 도메인 모델은 비즈니스 시스템의 규모 확장과 대량의 다양한 데이터들이 생성되는 빅데이터 환경, 그리고 데이터를 보다 스마트하게 처리하여 효율성 및 활용도를 높일 수 있는 방법으로 주목 받고 있다. 특히, 논리적인 단일 도메인 모델은 컴퓨팅 리소스의 유연함과 공유 경제에 의한 물리적으로 분할된 멀티 사이트의 데이터 처리를 위한 네트워크, 스토리지, 그리고 컴퓨팅 자원의 효율적 운영이 매우 중요하다. 기존의 Data Lake 프레임워크의 장점들을 기반으로 다양한 영역의 멀티 사이트들을 통합 및 데이터의 라이프 사이클을 관리하기 위한 DataLake 프레임워크의 Connected Data Architecture 개념과 기능들의 확장을 통한 다양한 응용 영역에 활용 가능한 CDA 기반 AI 서비스의 초안 설계 및 시나리오를 제안하고자 한다.

■ 중심어 : 데이터레이크; Abyss 스토리지 클러스터; 연결 데이터 아키텍처; AI 서비스

## Abstract

Single domain model like DataLake framework is in spotlight because it can improve data efficiency and process data smarter in big data environment, where large scaled business system generates huge amount of data. In particular, efficient operation of network, storage, and computing resources in logical single domain model is very important for physically partitioned multi-site data process. Based on the advantages of Data Lake framework, we define and extend the concept of Connected Data Architecture and functions of DataLake framework for integrating multiple sites in various domains and managing the lifecycle of data. Also, we propose the design of CDA-based AI service and utilization scenarios in various application domain.

■ keywords : DataLak; Abyss Storage Cluster; Connected Data Architecture; AI Service

## I. 서론

데이터는 기업 또는 조직 운영의 다양한 분야에서 중추적인 역할(Pivot-role)을 담당하면서 많은 기업과 조직들에게 중요성이 높아지고 있다. 기업과 조직의 가치는 데이터 중심(Data-centric)으로 변화하고 있으며, 데이터 중심은 점차적으로 모델 중심 비즈니스(Model-centric Business)로 가기 위한 중간 기착점 역할을 수행하고 있다. 여기서 모델은 곧 비즈니스이며, 모델 중심으로 가기 위한 전단계로 AI, 머신 러닝, 휴리스

틱 등의 기술들이 논의되고 있다. 딥러닝 등에 의한 인공지능은 만능 솔루션이 아니며, 한 모델이 모든 목적을 충족한다는 가정하는 경향을 피하는 것이 중요하다.

기업에서는 초창기에 데이터 저장소(Repository)에 자료를 저장 및 사용하는 단순한 방법으로 이용하였으나, 좀 더 고도화된 형태로 사용자의 의사 결정(Decision-making)에 도움을 주기 위하여, 조직 또는 기업 시스템의 데이터베이스에 저장 및 축적된 데이터를 공통의 형식으로 변환(Transformation)해서 관리(Management)하는 Data Warehouse를 이용하고 있다. 최근에는 빅데이터를 활용하는 사용 케이스들이 증가하고 있으

\* 정회원, 광주과학기술원 전기전자컴퓨터공학부

\*\* 정회원, 제노테크(주)

\*\*\* 정회원, 목포대학교 컴퓨터공학과

※ 본 연구는 과학기술정보통신부와 정보통신산업진흥원 및 광주정보문화산업진흥원의 “에너지신산업 SW융합클러스터조성사업(R&D, ITAS10101701100 70001000200200)”으로 수행된 연구결과입니다.

※ This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No.R7117-16-0218, Development of Automated SaaS Compatibility Techniques over Hybrid/Multisite Clouds).

접수일자 : 2018년 09월 06일

게재확정일 : 2019년 09월 27일

수정일자 : 2018년 09월 19일

교신저자 : 김종원, e-mail : jongwon@gist.ac.kr

며, 기업 또는 조직에서 처리되어야 하는 데이터들도 대량화되고 다양한 형태로 구성되어짐에 따라서 이를 좀 더 효율적으로 가공 및 처리할 수 있는 데이터 레이크 기술이 점차적으로 주목을 받고 있다. 여기서 멈추지 않고 비즈니스 인텔리전스를 유도하기 위한 기계 학습(Machine Learning) 및 인공지능(Artificial Intelligence) 등의 기술들이 적용되고 있는 상황이다. 기업 또는 조직의 비즈니스 시스템들에 의한 방대한 양의 데이터가 생성하며, 규모가 커지면서 서로 다른 이기종 시스템들에 존재하는 대량의 다양한 데이터를 보다 스마트하게 처리하기를 원하고 있다. 이를 위한 가장 기본적인 접근 방법 중 하나는 정보를 전사적 데이터화(Enterprise-wide Data)하여 데이터를 정확하게 표현하고 전체 비즈니스를 설명할 수 있는 논리적인 단일 도메인 모델(Single Logical Domain Model)을 구축하는 것이다. 이러한 일반적인 Data Lake 모델들은 생성되는 데이터를 캡처링(Capturing), 데이터 처리(Processing), 데이터 분석(Analyzing), 그리고 사용자 또는 데이터를 소비하는 시스템(Consuming systems)에 제공하여야 한다. Data Lake의 다양한 장점들을 획득하기 위해서는 7 가지의 Data Lake 요구 기능들(Data Governance & Data Lineage, Business Intelligence, Predictive Analysis, Information Traceability & Consistency, Historical Analysis to Derive Dimensional Data, Optimized Data Service, Lambda Architecture)을 어떤 구조로, 어떻게 동작 할 것인지에 대한 Data Lake의 구성요소를 정의하고, 확장성, 그리고 호환성 측면에서의 고려가 매우 중요하다[1].

본 논문에서는 DataLake 프레임워크의 장점들을 기반으로 다양한 응용 영역의 물리적으로 분할된 멀티 사이트들을 통합 및 데이터의 라이프 사이클을 관리하기 위한 Connected Data Architecture의 개념을 확장하며, 비즈니스 인텔리전스를 유도하기 위한 AI 서비스의 초안을 설계하고, AI 서비스를 활용하는 시나리오들을 제안한다.

## II. 관련 연구

### 1. Abyss Storage Cluster와 DataLake Framework

중소기업의 데이터센터를 위한 대용량 Abyss Storage Cluster는 Abyss Storage의 성능 향상을 위하여 스토리지의 디스크 매체별 성능 테스트와 네트워크 본딩(Bonding)에 의한 스토리지의 내부 네트워크의 가속화, 그리고 KOREN(Korea Advanced Research Network)과 TEIN(Trans-Eurasia Information Network) 네트워크를 이용한 국내 및 해외 네트워크의 트래픽 테스트를 완료한 상태이다[2, 3]. Abyss

Storage Cluster 기반의 DataLake 프레임워크는 다양한 요구 조건들을 고려하여 전사적 데이터 레이크(Enterprise-wide Data Lake)를 [그림 1]와 같이 정의 및 구성하였다[4].

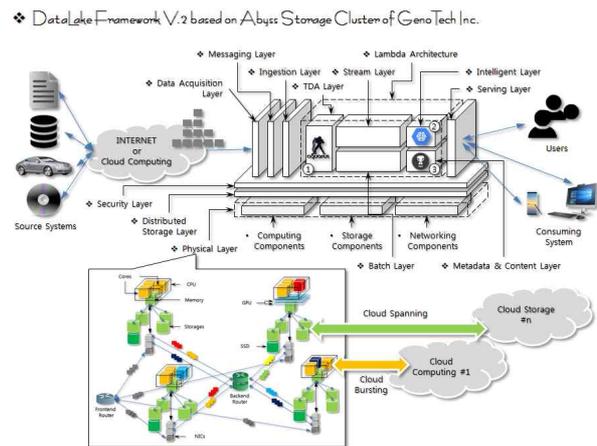


그림 1. DataLake Framework와 Abyss Storage Cluster 기반의 Cloud Bursting/Spanning

DataLake Framework의 최 하단에 위치한 Physical Layer는 DataLake의 논리적 기능(Logical functions)들을 지원하기 위한 물리적인 자원들(컴퓨팅, 스토리지, 그리고 네트워크 등)로 구성되며, 본 연구에서는 Abyss Storage Cluster가 실질적인 물리적인 컴퓨팅, 네트워크, 그리고 스토리지 자원들을 지원하게 된다. Physical Layer 위의 Distributed Storage Layer는 데이터 스트림(Data Stream) 및 이벤트(Event)에 대한 람다 아키텍처의 반응성(Reactivity)을 결정하게 된다. DataLake의 Security Layer는 다양한 보안 기능을 제공하며 각 계층의 구성 요소들 간의 식별(Identification), 인증(Authentication), 권한(Authorization), 네트워크 분리(Network Isolation), 데이터 방어(Data Protection), 그리고 감사(Auditing)/진단(Diagnose) 등을 지원한다. Data Acquisition Layer의 주요한 역할 중 하나는 데이터를 DataLake에서 추가적으로 처리 할 수 있는 메시지 변환 기능과 다양한 스키마를 수용할 수 있는 유연성(Flexibility)을 제공하며, 동시에 모든 변환된 데이터 메시지를 DataLake에 효율적으로 푸시(Push) 할 수 있는 연결 매커니즘을 제공한다.

Messaging Layer는 Data Acquisition Layer에 연결된 외부 시스템들 간의 멀티 연결들을 분리하는 계층이며 메시지 지속성(persistent)을 위한 메시지 전달을 보장하며, 메시지 지속성의 특성은 대개 스토리지 매체(Storage media)에서 지원한다. Ingestion Layer는 람다 아키텍처를 지원하기 위한 계층이며, 이 계층은 람다 아키텍처의 작업 모델로 데이터를 전달하는 속도 등을 제어하게 된다. 람다 아키텍처는 데이터의 일괄 처리

와 실시간 처리의 병합 문제를 해결할 수 있으며, 고성능(High-performance)의 분산 컴퓨팅 자원과 확장성으로 대규모 데이터 집합을 일괄적으로 거의 준 실시간(near-realtime)의 데이터 처리를 지원한다. 람다 아키텍처의 내부에는 TDA(Topology Data Analysis), Batch, Stream, Intelligent, 그리고 Meta-data & Content 등의 Layer들로 구성되며, DataLake Framework의 주요한 특성 중의 하나이다, 특히, 람다 아키텍처의 TDA Layer는 Topology data analysis[5]를 통한 원시 데이터(Raw data)를 모델링된 데이터(Modeled Data)로 변환하는 기능이 주요한 역할임과 동시에 다른 Data Lake 모델들과의 차별화되는 핵심 기능이다. 모델링된 데이터는 람다 아키텍처의 Serving Layer에서 제공 가능한 데이터 모델을 의미한다. 람다 아키텍처의 Intelligent Layer는 모델링된 고품질 데이터를 생성하기 위해 수집된 원시 데이터를 기반으로 머신 러닝(Machine Learning) 및 데이터 과학 처리를 Batch & Stream Layer에서 지원한다. 또한 모델링된 고품질 데이터를 Distributed Storage Layer에 저장되며, 사용자들과 다양한 소비 시스템들(Consumer Systems)을 위한 Service Layer에 제공하게 된다. 람다 아키텍처의 Meta-data & Content Layer는 DataLake 운영을 위한 생성된 메타데이터(Meta-data)의 저장/검색/관리 기능 등의 메타 데이터를 구성 및 관리하는 역할을 수행하며, DataLake를 구성하는 여러 Layer들의 다양한 기능들을 수행할 소프트웨어들을 가상화 기술에 의한 CI/CD/CD(Continuous Integration / Continuous Delivering / Continuous Deployment) 기능들을 제공하게 되며, 이러한 기능은 다른 Data Lake 모델들과의 차별화된 기능들을 제공하게 된다. 그리고 람다 아키텍처의 Meta-data & Content Layer에서 제공되는 클라우드 버스팅(Cloud Bursting)[6]과 클라우드 스패닝(Cloud Spanning)[7]은 애플리케이션의 작업 부하에 대한 원활한 운영과 컴퓨팅과 스토리지의 고가용성(High-availability)과 확장성을 보장한다. 마지막으로 DataLake의 Serving Layer는 DataLake로부터 사용자 또는 소비하는 애플리케이션들에게 데이터를 제공하는 역할을 수행하며, 특히 람다 아키텍처에서 생성된 데이터를 사용자 또는 소비하는 다양한 애플리케이션들에게 어떻게 전달 및 제공하는가가 중요하다. 모델링된 고품질 데이터는 다양한 이기종 시스템들 간의 다중 방법으로 전송이 가능하며, 데이터 서비스(Data Services)를 이용하여 빠른 데이터 전송이 이루어진다[8].

## 2. DataLake Framework의 CDA 개념

Hortonworks 회사의 Mark Haring의 인터넷 기고(起稿)문에서 Connected Data Architecture 개념이 소개되었으며,

Data Lake 개념은 데이터 웨어하우스의 진화의 결과물이며, Mark Haring은 서로 연결된 Data Pools 또는 Connected Data Architectures(CDA)에 대한 추가적인 사고(思考)가 필요하다고 언급하였다[9, 10].

DataLake 프레임워크에서 CDA 개념의 필요성과 역할로 논리적인 단일 도메인 모델을 구축하는 것이며, 이러한 일반적인 Data Lake 모델들은 데이터를 캡처링, 처리, 분석, 그리고 데이터를 소비하는 시스템들과 사용자들에 제공할 수 있는 전사적 데이터 레이크의 구축을 목표로 하며, 다양한 응용 영역에 데이터 전송을 위한 데이터 전송 서비스에 초점을 맞추고 설계되었다.

Abyss Storage Cluster 기반 DataLake Framework의 Connected Data Architecture 개념은 물리적으로 분할된 다양한 스토리지들(micro-Storage, Data Center, Cloud Storage 등)을 SDN(Software Defined Network) 기술에 의해서 논리적으로 하나로 묶을 수 있다는 개념이며 추상적으로 인터페이스(Interface)라 명명하며, 구현 측면에서는 [그림 2]와 같이 3가지 형태의 인터페이스를 정의 및 설계한다[11].

- DataLake와 Data Lake 간의 Interface #1
- DataLake와 Cloud Storage 간의 Interface #2
- DataLake와 micro-Storage 간의 Interface #3

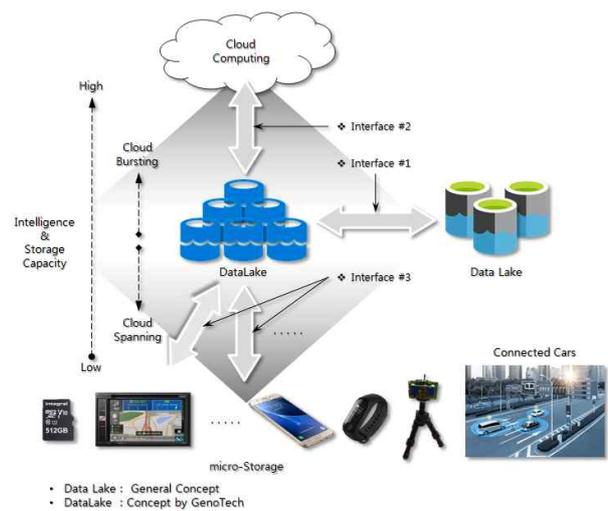


그림 2. Connected Data 개념의 인터페이스 정의

DataLake 프레임워크의 CDA 개념의 유스케이스로는 Interface 사례 별로 각각의 필요 기능과 제약 사항들을 기술한다. 위의 [그림 2]에서 DataLake를 중심으로 인터페이스(Interface) #1은 물리적 스토리지로 Abyss Storage Cluster 기반 DataLake 프레임워크와 다른 데이터 센터들 간의 SDN

기술을 통한 데이터 저장 공간의 확장 및 백업, 다중 복사, 그리고 이를 지원하기 위한 안전한 데이터 전송 및 보안을 위한 추가적인 컴퓨팅 자원의 지원이 필요하다.

CDA의 인터페이스 #2는 DataLake 프레임워크와 퍼블릭 클라우드의 스토리지 자원과의 SDN 기술을 통한 안전한 데이터 전송 및 데이터 저장 공간의 확장, 고도화된 서비스를 위한 컴퓨팅 자원의 지원, 그리고 보안이 필요하다. 특히 Interface #2는 데이터 공간 확보 측면보다는 클라우드 컴퓨팅 자원을 이용한 예측(Prediction) 및 Intelligence Analysis 등의 다양한 비즈니스 인텔리전스 서비스 지원에 의미가 있다.

마지막으로 CDA의 인터페이스 #3은 DataLake와 micro-Storage를 지원하는 다양한 IoT 디바이스들, 네비게이션, 커넥티드 카(Connected Car), 인포테인먼트 디바이스, 스마트폰 등과의 무선 네트워크 연결을 통한 다양한 IoT 디바이스들의 스토리지 공간을 제공하며, 추가적으로 비즈니스 인텔리전스 서비스를 위한 컴퓨팅 자원을 지원하게 되며, 가장 취약한 부분이 보안으로 예상된다.

### 3. 기계 학습 분야의 동향

4차 산업혁명을 주도할 기술은 빅데이터와 AI 라고 전문가들은 언급하고 있으며, [그림 3]은 Forrester Research[12]에서 2017년 1분기에 작성된 AI 기술의 TechRader를 나타낸 것이며, 특히 기계 학습 분야는 NIPS 2016[13]에서 [그림 4]와 같이 산업 동향을 예측하고 있다. 특히, 머신 러닝 분야에서 지도 학습(Supervised Learning) 분야와 비지도 학습(Unsupervised Learning) 분야의 간극을 전달 학습(Transfer Learning)[14]이 어느 정도 채워주고 있는 상황이다.

특히 인공지능의 내장형 AI(Embedded AI)[15] 기술은 Smart Objects /Natural Language Processing/Personalization 등 개인맞춤형 서비스에 대한 관심이 높아지고 있으며, 향후 사업화 가능성이 높은 부분도 개인화/맞춤형/제안형 제품 위주를 지원하게 될 것으로 예상된다. 상업적 측면에서 내장형 인공지능 서비스로서, 헬스케어, 웨어러블 디바이스, IoT 디바이스 등 저사양 저전력 디바이스에 적용이 예상된다. 특히 소형, 저전력, 부착형 디바이스에 AI 기술을 접목하여, 사용자의 패턴분석을 통한 사용자 환경인지, 사용자 상태인지 등을 통해 맞춤형 서비스가 필요한 헬스케어, 웨어러블 디바이스 산업, 인포테인먼트(Infotainment)[16] 등에 활용될 수 있다. 기술적 측면에서 하드웨어 사양이 낮고 저전력을 요구하는 임베디드 시스템 내 적합한 인공지능 솔루션 개발을 위해서는 초소형 학습기능 하드웨어 개발을 통해 서버로 통합되는 데이터양을 줄이고 응용제품 데이터의 특화된 최적의 학습 및 데이터 분류가 가능하도록 개발될 것이며, 저사양의

CPU 코어/ 병렬처리 기술/메모리 활용을 통한 학습 데이터 처리 및 서비스에 맞는 데이터 전처리 기술 지원이 필요하다.

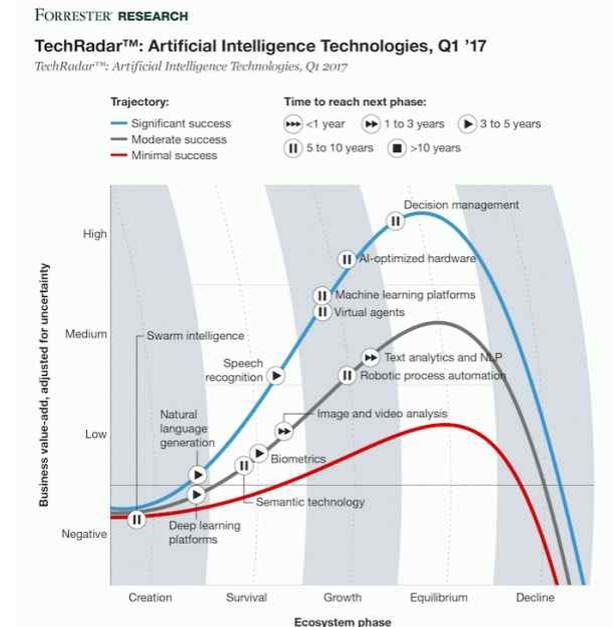


그림 3. Forrester Research의 AI 기술 동향 예측

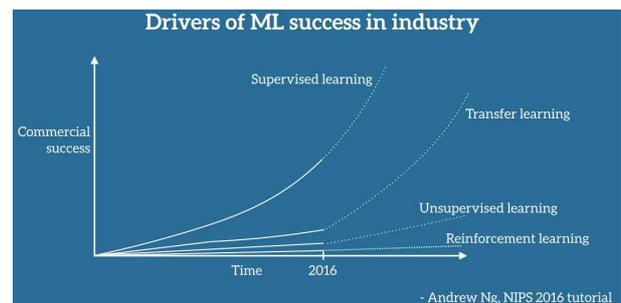


그림 4. 기계 학습 분야의 산업 동향 예측

### III. CDA 개념 확장을 통한 AI 서비스의 초안 설계 및 시나리오 제안

본 연구에서 제안하는 DataLake Framework의 CDA 개념에서 컴퓨팅 자원의 확장을 통한 AI 서비스 개념의 초안을 [그림 5]와 같이 제안하며, 이를 실체화하기 위한 시나리오 4가지를 서술한다.

CDA 개념을 확장한 AI 서비스의 설계는 [그림 2]의 기존 CDA의 인터페이스를 기반으로 사용자 또는 데이터를 소비하는 시스템의 컴퓨팅 환경과 네트워크, 컴퓨팅, 그리고 스토리지 자원의 지원 측면에 초점을 맞추어 비즈니스 인텔리전스를 유

도하기 위한 AI 서비스를 설계하였다.

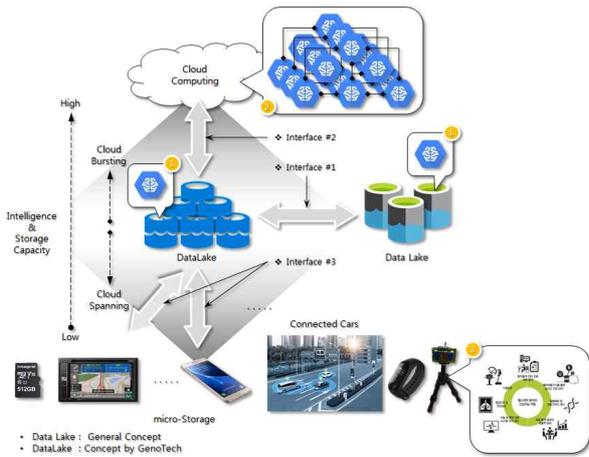


그림 5. CDA의 AI 서비스 개념

### 1. CDA 기반 AI 서비스의 시나리오 #1

CDA 인터페이스의 AI 서비스의 시나리오 #1은 DataLake를 중심으로 다른 Data Lake와의 협업을 통한 AI 서비스를 위한 AI 기본 3 단계(수집 단계, 학습 단계, 그리고 적용 단계)들을 공유 및 확장성을 지원하게 되며, [그림 5]의 ①과 같이 나타낼 수 있다. 수집 단계에서는 학습 데이터의 수집/저장/관리를 위한 DataLake의 물리적 자원의 한계성을 SDI(Software Defined Infrastructure) 기술에 의한 공유 및 확장성을 확보하게 된다. 또한, 학습 단계에서는 SDI 기술과 가상화 기술에 의한 AI 애플라이언스를 통해 학습을 병렬로 처리하게 된다. 또한 적용 단계에서는 학습된 결과를 공유 및 AI 서비스를 제공한다.

### 2. CDA 기반 AI 서비스의 시나리오 #2

CDA 인터페이스의 AI 서비스의 시나리오 #2은 [그림 5]의 ②와 같이 DataLake를 중심으로 퍼블릭 클라우드를 통한 스토리지 및 컴퓨팅 자원의 확장성을 취하게 되며 AI 서비스를 위한 수집 단계와 학습 단계 영역에서 확장성을 지원한다. 수집 단계에서는 학습 데이터의 수집/저장/관리를 위한 DataLake의 물리적 자원의 한계성을 극복하기 위하여 퍼블릭 클라우드의 스토리지로 확장성을 확보하며, 학습 단계에서는 퍼블릭 클라우드의 강력한 컴퓨팅 자원의 지원을 통한 AI 애플라이언스의 학습 및 학습 시간을 단축이 가능하게 된다.

### 3. CDA 기반 AI 서비스의 시나리오 #3

CDA 인터페이스의 AI 서비스의 시나리오 #3은 DataLake를 중심으로 다양한 IoT 애플라이언스의 프라이빗 특성을 가상화 기술로 지원하며, 다양한 IoT 디바이스의 micro-Storage과 저사양 컴퓨팅의 특성들에 의하여 AI 서비스의 적용 단계만을 지원하며, [그림 5]의 ③과 같이 표현하며, AI 서비스의 수집 단계와 학습 단계는 DataLake에서 지원하게 된다.

### 4. CDA 기반 AI 서비스의 시나리오 #4

CDA 인터페이스의 AI 서비스의 시나리오 #4는 시나리오 #2와 시나리오 #3을 통합한 시나리오이며, DataLake를 중심으로 퍼블릭 클라우드를 통한 스토리지 및 컴퓨팅 자원의 확장하며, AI 서비스를 위한 수집 단계와 학습 단계를 수행하게 된다. 학습이 완료되면, AI 서비스의 적용 단계에서는 DataLake에서 다양한 IoT 애플라이언스들에게 AI 서비스를 제공하게 되며, AI 서비스의 적용 단계 이전에 전이 학습이 수행되어야 한다.

### 5. CDA 기반 AI 서비스의 유효성

CDA 기반 AI 서비스를 제공하기 위한 플랫폼 프로토타입으로 파이썬 기반의 MXNet[17]을 이용하였으며, 시나리오 #3에 적용하기 위한 유효성과 객체 탐지의 가능성을 검증하였다. MXNet을 이용한 시나리오 #3에 적용하기 위한 시물레이션 결과는 [그림 6]과 같이 손 글씨의 패턴인식에 90%의 탐지율을 보였으며, [그림 7]은 MXNet의 딥 러닝을 이용한 이미지 객체 탐지의 가능성을 확인하였다.

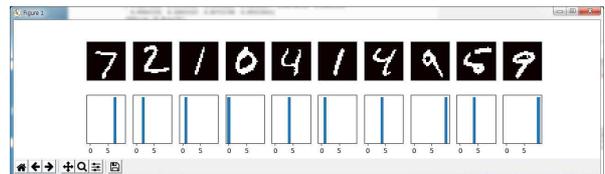


그림 6. MXNet을 이용한 손 글씨의 패턴인식 결과

### 3. CDA 기반 AI 서비스의 시나리오 #3

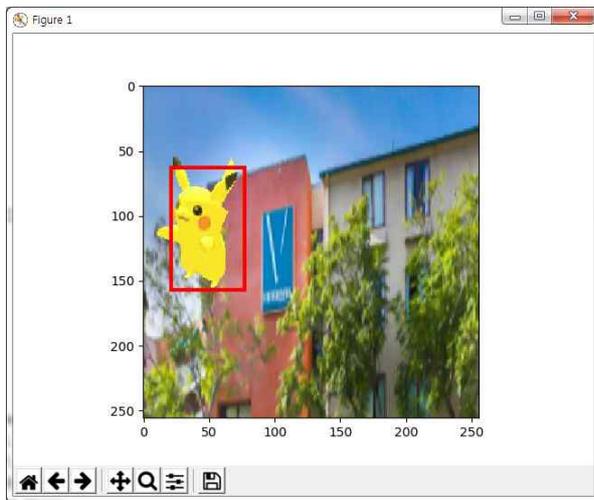


그림 7. MXNet의 딥 러닝을 이용한 이미지 객체 탐지

CDA 기반 AI 서비스를 제공하기 위한 플랫폼 프로토타입으로 MXNet은 Gluon 라이브러리에 의한 학습 속도를 그대로 유지하면서 손쉽게 딥 러닝 모델의 프로토타입을 구축, 교육 및 배포할 수 있는 상위 수준의 인터페이스를 제공하며, 특히 딥 러닝 워크로드는 선형에 가까운 확장성으로 여러 GPU로 분산될 수 있는 특징을 제공한다. 즉, 매우 규모가 큰 프로젝트도 상대적으로 짧은 시간 내에 처리할 수 있는 특징을 갖고 있다. 또한, IoT(Internet of Things) 및 엣지(Edge) 용도로 클라우드에서 여러 GPU 교육 및 복잡한 모델 배포를 처리하는 것 외에도 MXNet은 Raspberry Pi, 스마트폰 또는 노트북과 같은 저출력 엣지 디바이스에서 실행되어 데이터를 원격으로 처리할 수 있는 간단한 신경망 모델 표현의 생성이 가능하다. 또한, 클러스터 내 GPU 수에 따라 규모 조정도 자동으로 수행되며, 개발자는 서버리스 및 배치 기반 추론을 실행하여 시간을 절약하고 생산성을 높일 수 있는 특징을 제공한다.

#### IV. 결론

최근 데이터는 기업 운영의 중추적인 역할과 중요성이 높아지고 있으며, 점차적으로 기업의 가치는 데이터 중심으로 이동하고 있으며, 종착역으로 모델 중심 비즈니스를 지향하고 있다. Abyss Storage Cluster 기반 DataLake 프레임워크의 Connected Data Architecture 개념은 물리적으로 분할된 다양한 스토리지들을 소프트웨어에 의해 논리적으로 묶을 수 있다는 개념을 수립하며, 한 단계 더 나아가 비즈니스 인텔리전스를 유도하기 위한 AI 서비스의 초안을 설계하였다. 더불어 CDA 기반 AI 서비스를 실체화하기 위한 시나리오 4가지를 제안하였다.

#### REFERENCES

- [1] Tomcy John, Pankaj Misra, "Data Lake for Enterprises - Leveraging Lambda Architecture for Building Enterprise Data Lake," Packt Publishing, May, 2017.
- [2] 차윤석 외 4인, "Abyss Storage의 Disk 타입에 의한 Ceph RADOS의 Benchmarking," 2017 한국통신학회 동계학술대회, 1271-1273쪽, 하이원리조트, 대한민국, 2018년 6월
- [3] 차병래 외 4인, "대용량 Abyss Storage의 KOREN 네트워크 기반 국내 및 해외 실증 테스트," 스마트미디어저널, 제6권, 제1호, 9-15쪽, 2017년 3월
- [4] 차병래 외 3인, "Abyss Storage Cluster 기반의 DataLake Framework의 설계," 스마트미디어저널, 제7권, 제1호, 9-15쪽, 2018년 3월
- [5] 차병래 외 4인, "Idea Sketch to Improvement Image Learning based on Machine Learning using Topology Theory," SMA 2017, pp.139-141, Boracay, Philippine, Dec., 2017.
- [6] Cloud Bursting(2017). <http://searchcloudcomputing.techtarget.com/definition/cloud-bursting> (accessed Sep., 7, 2018).
- [7] Cloud Spanning(2014). <http://searchcloudcomputing.techtarget.com/definition/cloud-spanning>(accessed Sep., 7, 2018).
- [8] 차병래 외 4인, "Abyss Storage Cluster 기반 DataLake Framework의 설계," 스마트미디어저널, 제7권, 제1호, 9-15쪽, 2018년 3월
- [9] Mark Harring, "Connected Data Ponds: The evolution of Data Lakes," Hortonworks, Sept. 08, 2016.
- [10] 차병래 외 3인, "Abyss Storage Cluster 기반 DataLake Framework의 Connected Data Architecture 개념 설계," 2018 한국정보기술학회 하계공동학술대회, 4-7쪽, 조선대학교, 대한민국, 2018년 6월 8일
- [11] 차병래 외 4인, "Abyss Storage Cluster 기반 DataLake Framework의 Connected Data Architecture 개념 설계 및 검증," 스마트미디어저널, 제7권, 제3호, 57-63쪽, 2018년 9월
- [12] Forrester Research(2018), <https://www.forrester.com/> (accessed Sep., 7, 2018).
- [13] NIPS 2016. <https://nips.cc/Conferences/2016> (accessed Sep., 7, 2018).
- [14] Transfer Learning(2018), [https://en.wikipedia.org/wiki/Transfer\\_learning](https://en.wikipedia.org/wiki/Transfer_learning) (accessed Oct., 12, 2018).
- [15] Embedded AI(2018), <https://blog.g2crowd.com/blog/trends/artificial-inte>

lligence/2018-ai/embedded/ (accessed Oct., 12, 2018).

- [16] Infotainment(2018), <https://en.wikipedia.org/wiki/Infotainment> accessed Oct., 12, 2018).
- [17] MXNet(2017), <https://mxnet.apache.org/> (accessed Dec., 9, 2017).

저 자 소 개



차병래

2004년 목포대학교 대학원 컴퓨터공학과 졸업(공학박사)  
 2005년 호남대학교 컴퓨터공학과 전임 강사  
 2009년 ~ 현재 광주과학기술원 전기전자컴퓨터공학부 연구조교수

2012년 ~ 현재 제노테크(주) 대표이사  
 <주관심분야: 정보보안, IDS, Neural Network, Cloud Computing, VoIP, NFC, 대용량 스토리지 기술 등>



박 선

2007년 인하대학교 컴퓨터정보공학과 공학박사  
 2008년 호남대학교 컴퓨터공학과 전임 강사  
 2010년 전북대학교 인력양성사업단 박사후 과정

2010년 목포대학교 정보산업연구소 연구전임교수  
 2013년 ~ 현재 광주과학기술원 NetCS연구실 연구교수  
 <주관심분야: 정보검색, 데이터마이닝, 해양IT정보융합, 클라우드 컴퓨팅, IoT, 스토리지 시스템>



오수열(증신회원)

1981년 전남대학교 공학사  
 1985년 조선대학교 전산학석사  
 2004년 전남대학교 전산학박사수료  
 2005년 호주 타스마니아대학 객원교수

현재 목포대학교 컴퓨터공학과 교수  
 <주관심분야: 프로젝트관리, 소프트웨어 품질, 빅데이터, 공학교육>



김종원

1997년 University of Southern California 연구 조교수  
 1999년 Technology Consultant for VProtect Systems Inc.  
 2000년 Technology Consultant for Southern California Division

of InterVideo Inc.  
 2001년 광주과학기술원 전기전자컴퓨터공학부 교수  
 2008년 ~ 현재 광주과학기술원 전기전자컴퓨터공학부 교수

<주관심분야: Future Internet, SDN & NFV, SDI>