

HELLO,

# 헬로, 핀테크!

핀테크 기반기술

FINTECH



한국핀테크지원센터  
Fintech Center Korea

## 발간사

핀테크(Fintech)는 말 그대로 금융(Finance)과 기술(Technology)의 만남입니다. 핀테크는 기술을 통해 우리가 이용하는 금융을 더 간편하게 하고, 손안의 금융으로 만들어 줍니다. 특히, 최근에는 빅데이터, AI, 클라우드, 블록체인 등 디지털 신기술이 금융분야에 적극적으로 도입되면서 금융의 디지털화, 플랫폼화, 탈중개화 등 금융혁신이 급격히 진행되고 있습니다. 이에 따라 세계 각국은 핀테크를 금융산업의 혁신과 융합을 촉진하는 새로운 성장동력으로 삼고 국가적인 지원정책을 경쟁적으로 펼치고 있습니다.

우리나라도 핀테크 산업 육성을 위해 금융규제 샌드박스 제도 도입·운영, 신용정보법 등 데이터 3법 개정, 오픈 API 구축, 한국핀테크지원센터 설립 등 적극적인 혁신금융정책을 추진하여 디지털 금융서비스가 다른 나라와 비교해 빠른 속도로 시장에 안착되고 있으며, 이를 통해 소비자의 금융 편의성이 제고되고 금융권의 디지털 전환(Digital Transformation)이 빠르게 진행되고 있습니다. 한국핀테크지원센터는 이러한 정부의 노력을 뒷받침하고 핀테크 생태계 구축 및 성장 지원을 위해 핀테크 기업에 대한 멘토링, 교육, 입주공간, 금융규제 샌드박스, 해외진출, 투자유치 등을 지원하고 있으며, 특히 금년에는 새로운 아이디어를 출시 전에 금융권의 실제 데이터를 바탕으로 테스트해볼 수 있는 D-테스트베드, 금융사와 핀테크 기업의 협력을 위한 금융사-핀테크 협력플랫폼, 이미 해외진출을 추진한 기업들의 경험을 공유하는 해외진출 플랫폼 등을 통해 혁신적인 아이디어를 가진 창업자들이 창업과 스케일업을 거쳐 유니콘으로 성장할 수 있도록 돕고 있습니다. 앞으로도 한국핀테크지원센터는 금융당국과 함께 기술기반 핀테크 기업을 발굴·육성하여 국내 디지털 금융혁신을 촉진하고 글로벌 유니콘 핀테크 기업으로 성장할 수 있도록 적극 지원해 나가겠습니다.

한편, 핀테크 분야는 앞으로 인력 수요는 크게 늘어나는데 비해, 인력 공급이 이를 따라가지 못해 전반적으로 인력 부족 현상이 심화될 것으로 전망됩니다. 따라서 핀테크를 이해하는데 필요한 기술적 측면과 핀테크 서비스의 개발·출시



등을 위해 알아야 할 법령 등 각종 금융관련 규제 등을 종합적으로 다룬 핀테크 교재의 필요성이 커지고 있는 상황입니다.

이에 지난 2020년, 금융위원회와 한국핀테크지원센터는 「헬로, 핀테크!」 교재 6종을 발간하여 핀테크 산업에 관심있는 일반인부터 핀테크 분야 재직자까지 누구나 핀테크 입문부터 지급결제·송금, 금융플랫폼·금융데이터, 자산관리·보험, 보안인증·블록체인, 개인신용정보 관리·활용 등 분야별 지식까지 습득할 수 있도록 하였습니다.

본 교재는 발간 이후 핀테크 기업, 금융회사, 대학 등 다양한 곳에서 핀테크 지식을 습득하는데 활용되고 있습니다. 최근 1년 사이에 핀테크 분야에도 많은 변화가 있었고, 이 변화를 교재에 수록했으면 하는 요청들이 많아 이번에 새롭게 개정판을 발간하게 되었습니다.

이번 「헬로, 핀테크!」 개정판은 빠르게 변화하는 핀테크 분야 규제와 시장 현황 등 내·외부 변화에 맞추어 그동안 제·개정된 법률, 기술 등 다양한 변화를 최대한 수록하였습니다. 이와 함께, 핀테크 분야에서 꼭 필요한 기반기술인 AI, 빅데이터, 클라우드, IoT 등에 대한 기초 지식을 습득하고 실습해 볼 수 있도록 ‘핀테크 기반기술’ 편을 새로이 발간하였습니다.

시중에 핀테크를 다룬 많은 책들이 있습니다만, 이 교재는 핀테크 분야의 현황과 각종 금융관련 규제·제도의 변화를 종합적으로 정리한 교재라는 점을 약속드리며, 모쪼록 본 교재가 핀테크 기업과 핀테크에 관심있는 모든 분들께 잘 활용되고, 우리나라 핀테크 분야의 경쟁력을 높이고 핀테크의 혁신과 성장에 이바지할 수 있기를 기원합니다.

2021년 11월  
한국핀테크지원센터 이사장  
**변영한**

## 추천사

IT 기술의 발전은 금융의 많은 것을 바꿔놓고 있습니다. 금융회사를 직접 방문하지 않고 스마트폰 앱을 이용하여 간편하게 저축·송금·투자를 진행할 수 있으며, 인공지능(AI) 기술이 도입된 로보어드바이저를 통해 금융상품 추천을 받고, 현금·카드가 없어도 스마트폰에 내장된 결제모듈을 이용해 언제 어디서든 편리하게 물건을 구매할 수 있게 되었습니다.

이처럼 핀테크는 이미 우리 생활 깊숙이 자리 잡아 개인이 원하는 맞춤형 금융서비스와 높은 편의성을 제공하고 있습니다. 금융소비자들은 금융서비스 이용을 위한 비용을 절감할 수 있고, 언제 어디서나 서비스를 이용할 수 있기에 시·공간적인 측면에서도 혜택을 누릴 수 있습니다. 여기에 중소기업 및 소상공인에 특화된 금융서비스를 개발하기도 용이하고, 저소득층, 고령층 등 금융 사각 지대에도 보다 다양한 기회를 제공하는 포용적 금융을 실천할 수 있습니다.

이에 정부에서도 핀테크 산업을 지원·육성하기 위하여 혁신금융서비스에 대한 금융규제 샌드박스 제도를 내실화하는 등 제도적 지원을 아끼지 않고 있으며, 금융회사들도 빠르게 변화하는 금융소비자의 니즈를 충족시키기 위하여 ‘디지털 트랜스포메이션’을 통한 새로운 조직으로 변화를 추구하고, 핀테크를 활용한 금융서비스 개발에도 누구보다 앞장서고 있습니다. 또한 핀테크 스타트업 기업들은 새로운 디지털기술을 도입하여 기존에 없던 금융서비스를 시도하며 금융회사들과 협업·경쟁을 통해 금융서비스의 고도화를 가속화하고 있습니다.

핀테크 산업은 다양한 관계자들의 협업·경쟁에 기반한 기술의 융합을 추구하고 있으며, 이를 지속적으로 발전시키기 위해서는 전문인력을 양성하고 관련 지식을 꾸준히 보급할 수 있는 생태계 환경조성이 무엇보다 필요합니다. 그런 의미에서 금융위원회와 한국핀테크지원센터에서 핀테크 특화 전문도서를 지속적으로 발간하는 것은 매우 바람직한 일이라 생각합니다.

그동안 한국금융연수원은 금융 DT 아카데미를 통해 금융인의 디지털 금융 및 핀테크 역량강화를 위한 전문가 육성에 선도적인 역할을 수행해 왔습니다. 이러한 전문성을 바탕으로 우리 원이 지난 해 「헬로, 핀테크!」 6종 도서 발간에 이어 금년에도 최신정보를 반영한 6종 도서의 개정판 발간과 함께, 최근 주목받고 있는 빅데이터와 인공지능(AI), 클라우드, 사물인터넷(IoT) 등 핀테크 기반기술 관련 도서 1종의 추가 신규 발간 참여를 통해 우리나라 핀테크 산업 발전에 작게나마 기여할 수 있게 되어 매우 뜻깊게 생각합니다.

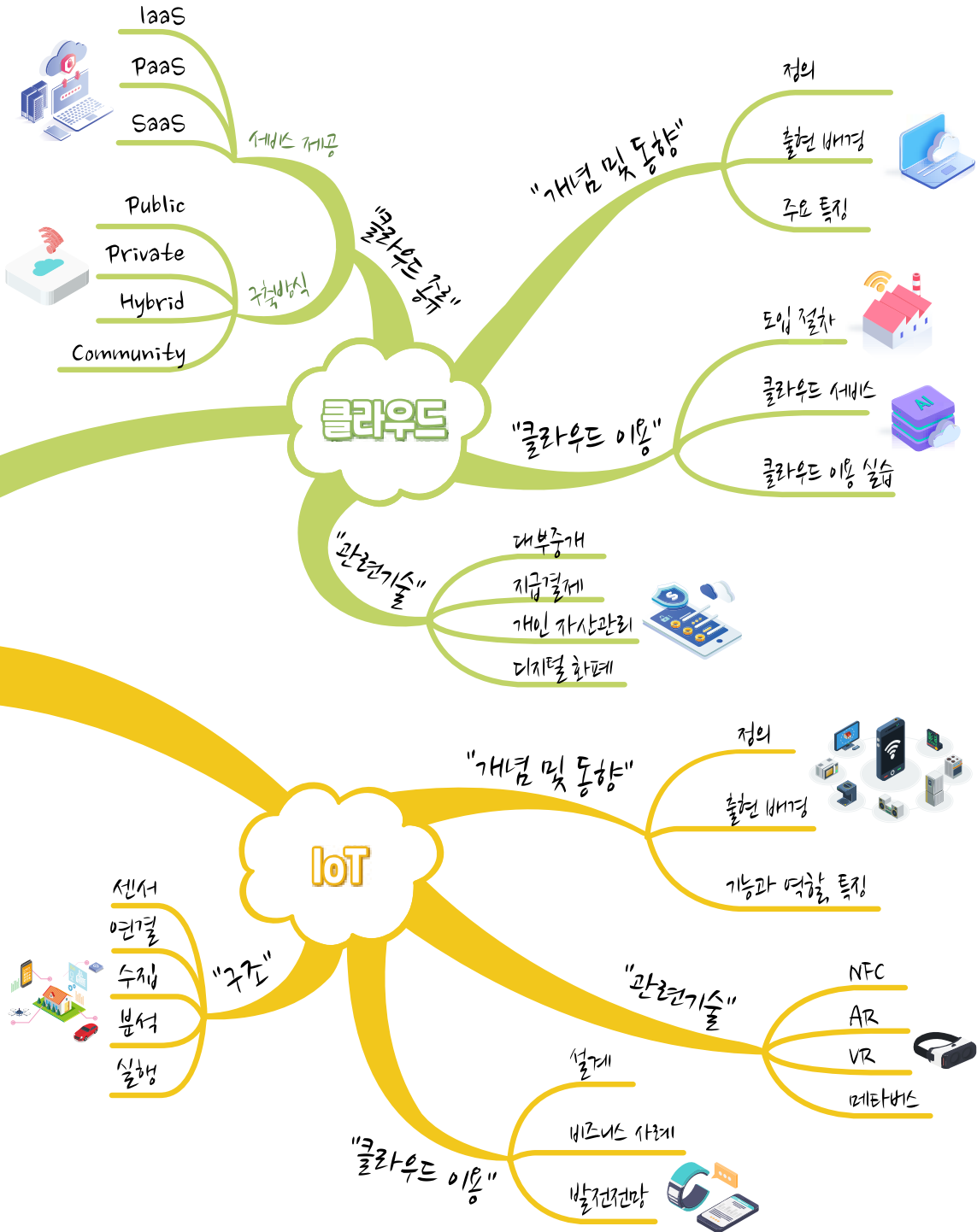
아무쪼록 본 도서가 현재 금융회사나 핀테크 업체에 근무하는 분들은 물론, 앞으로 해당 분야 취업이나 창업을 준비하는 분들에게도 많은 도움이 될 수 있기를 바랍니다. 그리고 무엇보다 핀테크 산업에 대한 국민적인 관심과 이해를 높이는 데 좋은 길잡이가 되기를 희망합니다.

2021년 11월  
한국금융연수원 원장  
**서태종**

HELLO,  
FINTECH

# 헬로, 핀테크!(핀테크 기반기술) 학습맵





# CONTENTS



## 제1장 빅데이터와 인공지능

제1절 처음 만나는 빅데이터와 인공지능	14
제2절 빅데이터 바로알기	20
제3절 빅데이터 처리 및 운영	27
제4절 인공지능의 이해	39
제5절 인공지능 학습법의 분류	47
제6절 머신러닝과 딥러닝 바로 알기	69
제7절 빅데이터와 인공지능의 미래 전망	83

## 제2장 클라우드

제1절 왜 클라우드인가? (클라우드 소개 및 효과)	100
제2절 클라우드란? (클라우드의 이해)	112
제3절 클라우드를 사용하려면? (클라우드 도입 과정)	130
제4절 클라우드 사례 (사례를 통한 클라우드의 이해)	146
제5절 클라우드와 핀테크	165
제6절 주요 클라우드 서비스 이해	173
제7절 실제로 클라우드 사용해 보기	213

## 제3장 IoT

제1절 처음 만나는 IoT	240
제2절 IoT의 탄생	243
제3절 IoT의 기능과 역할, 구조, 주요 특징	251
제4절 IoT 설계	264
제5절 IoT와 핀테크	291
제6절 IoT 미래 전망	301
제7절 IoT와 좀 더 친해지기	310

헬로, 핀테크!(핀테크 기반기술)



HELLO, FINTECH!



# 1 장

## 빅데이터와 인공지능

제1절 처음 만나는 빅데이터와 인공지능

제2절 빅데이터 바로알기

제3절 빅데이터 처리 및 운영

제4절 인공지능의 이해

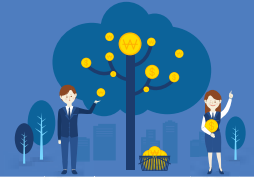
제5절 인공지능 학습법의 분류

제6절 머신러닝과 딥러닝 바로 알기

제7절 빅데이터와 인공지능의 미래 전망

1 장

# 빅데이터와 인공지능



## 💡 학습목표

- ① 빅데이터와 인공지능이 주목받는 배경을 설명할 수 있다.
- ② 빅데이터가 가진 성질을 설명할 수 있다.
- ③ 빅데이터의 주요 특징과 부가적 특징을 설명할 수 있다.
- ④ 인공지능의 정의를 말할 수 있다.
- ⑤ 지도학습과 비지도 학습이 무엇인지 설명할 수 있다.
- ⑥ 강화학습이 무엇인지 설명할 수 있다.
- ⑦ 머신러닝이 무엇인지 설명할 수 있다.
- ⑧ 딥러닝이 무엇인지 설명할 수 있다.
- ⑨ 전이학습이 무엇인지 설명할 수 있다.
- ⑩ 연합학습이 무엇인지 설명할 수 있다.
- ⑪ 인공지능으로 할 수 있는 것과 없는 것을 구분할 수 있다.
- ⑫ 인공지능의 비즈니스 사례에 대해 설명할 수 있다.

## 💡 학습개요

하드웨어의 비약적인 발전은 우리 삶의 많은 부분을 변화시켰다. 저렴해진 컴퓨터 처리 장치는 다양한 기기를 양산했고, 지구에는 이러한 기기들이 생산하는 방대한 자료가 끊임없이 쏟아지고 있으며, 대부분은 소셜미디어 등을 통해 정형화되지 않은 형태로 만들어 지며, 우리는 이를 빅데이터라 부른다. 한편, 과거 컴퓨터 계산 자원의 제약으로 인해 크게 주목받지 못하던 딥러닝 기술은 하드웨어의 발전과 함께 빅데이터의 활용을 통해 이제 데이터에 비례한 더 비약적인 성능을 얻을 수 있는 토대를 마련했다. 우리는 이를 흔히 인공지능 기술이라 부른다.

빅데이터는 생성되는 빈도와 소스 그리고 그 형태에 있어서 일반적 데이터와 크게 차이가 난다. 이로 인해 발생하는 빅데이터가 가진 공통적인 요소 다섯 가지를 살펴보고, 금융에 있어 빅데이터의 활용에 대해 살펴본다.

또한 책에서는 인공지능의 기본적인 세 가지 학습 방법인 지도학습, 비지도 학습, 강화 학습에 대해 자세히 알아본다. 또한 일반적인 인공지능 기법인 판별적 기술과 달리 무에서 유를 창출하는 생성적 기법도 살펴보고, 이와 관련하여 머신러닝과 딥러닝에 대해 살펴본 다음 인공지능이 사용되는 여러 비즈니스 사례를 차례로 살펴본다.

마지막으로는 인공지능에 관한 여러 개념들인 전이학습, 연합학습, 자동 머신러닝, 해석가능 인공지능의 개념을 차례로 정리하고 끝으로 인공지능으로 할 수 있는 것과 없는 것을 구분할 수 있는 방법을 설명한다.

### 용어해설

#### ① 사이버네틱스(Cybernetics)

2차 대전 후 미국의 수학자 위너에 의해 제창된 학문으로서 기술장치, 생물 유기체, 인간의 여러 조직의 공통된 각종 특징을 찾아내고자 한 학문

#### ② 오토마타(Automata)

스스로 움직이는 기계를 칭하는 말

#### ② 관계형 데이터 베이스

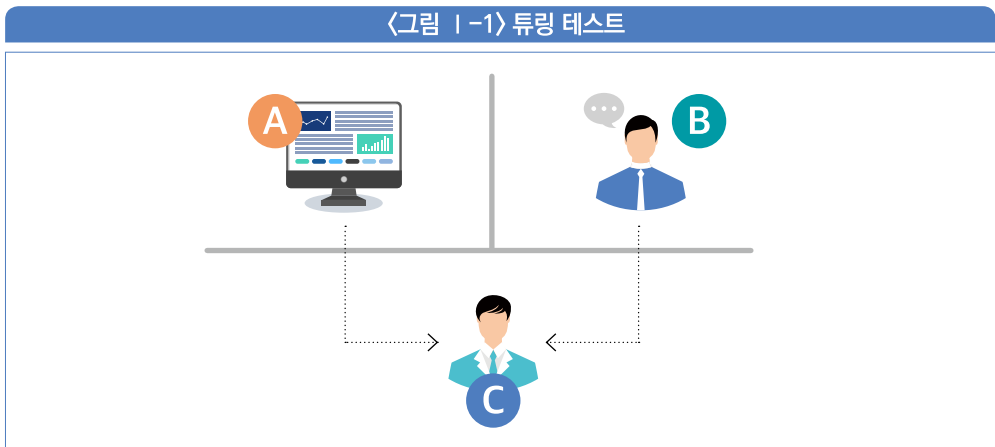
다수의 원활한 사용을 위해 관리할 목적으로 모아진 데이터의 집합을 데이터베이스라고 하며, 그중에서 각 객체들 사이의 관계를 표 형식으로 정의 해 둔 것을 관계형 데이터베이스라고 함.

## 1 다시 주목받는 인공지능

인공지능(Artificial Intelligence)이라는 말이 처음으로 생겨난 것은 1956년으로 알려져 있다. 그전에는 인공지능 대신 “생각하는 기계(Thinking Machine)”, “사이버네틱스(Cybernetics)”, “오토마타(Automata)” 등 서로 다른 다양한 용어로 불렸으며, 막연히 사람과 유사한 기계를 만드는 기술 정도로만 묘사되었을 뿐 그 정확한 정의는 없었다.

사람과 유사한 기계에 대한 개략적이지만 명확한 가이드라인은 영국의 천재 수학자 앨런 튜링(Alan Turing)이 1950년에 처음으로 제시하였다.

〈그림 1-1〉 튜링 테스트



〈그림 1-1〉은 앨런 튜링이 제시한 튜링 테스트의 기본 개념도를 보여주고 있다. 튜링은 이 테스트를 통과한 기계는 ‘생각하는 기계’로 간주할 수 있다고 주장했으며, 이를 판별하는

기준으로 튜링 테스트를 제시했다. 테스트는 이렇게 진행된다. 안이 보이지 않는 두 개의 방이 있다. 이 중 하나의 방에는 컴퓨터가 들어있고 다른 하나의 방에는 실제 사람이 들어가 있다. 이때 방 외부에 있는 사람이 방안에 있는 그 무엇과 대화를 나누었을 때, 그것이 사람인지 기계인지 구분하지 못하게 될 때, 비로소 “생각하는 기계”가 구현된 것으로 간주할 수 있다고 제시한 것이다. 이 때문에 초창기에 많은 연구는 이 생각하는 기계의 구현에 집중되었고, 여러 특화된 컴퓨터 언어가 등장하기 시작했다. 1958년 등장한 LISP(LISt Processor)이라는 언어는 두 번째로 오래된 고급 컴퓨터 언어인데 접두사(prefix)처리에 특화된 기능으로 인해 당시 인공지능 구현에 있어 가장 주목받던 언어이기도 했다.

## 1-1 인공지능 용어의 탄생

‘생각하는 기계’가 대대적으로 주목을 받기 시작한 곳은 1956년 여름 일단의 학자들이 다트머스(Dartmouth) 대학에서 개최한 워크숍에서였다. 그 워크숍의 명칭은 바로 “Dartmouth Summer Research Project on Artificial Intelligence” 였으며, 이 워크숍에서 최초로 인공지능(Artificial Intelligence)이라는 용어가 세상에 등장한 것이다.

당시 인공지능에 들어 있는 ‘지능’이라는 용어에 많은 사람들이 현혹되었고 이러한 점은 다소 많이 과장되어, 마치 금방이라도 사람과 구분할 수 없을 정도의 성능을 가진 기계가 등장할 것처럼 한동안 세간에 알려지기도 했다. 그러나 이러한 과도한 기대는 곧 크나큰 실망을 낳게 된다.

## 1-2 인공지능 빙하기

과도한 기대가 낳은 실망으로 인해 모두 두 차례에 걸쳐 아무도 인공지능을 주목하지 않고 투자도 하지 않는 소위 ‘인공지능의 빙하기’가 도래한다. 투자자들의 높은 기대와 관심과는 달리 아무것도 보여주지 못하는 인공지능 기술은 그저 허황된 사기에 지나지 않게 느껴진 것이다.

두 번의 빙하기는 각각 1974년부터 1980년까지의 약 6년간과 1987년과 1993년까지의 약 6년간이다. 인공지능에 대한 보다 자세한 역사는 제4절에서 다시 살펴보기로 하자.

### 1-3 딥러닝 기술

과거의 컴퓨터는 지금과 비교할 때 그 연산능력이 현저히 떨어졌고, 지금의 딥러닝 기법을 사용하는 것은 불가능했다. 당시 대부분의 전통적 인공지능 기법은 연산 자원이 비교적 적게 소모되는 형태 즉, 대체로 사람이 많이 개입되어 여러 가지 가정을 한 후 데이터를 통해 검정하는 통계와 유사한 방식이었고 이 때문에 데이터를 조금 더 많이 확보한다고 하더라도 얻을 수 있는 성능향상은 매우 제한적이었다. 대부분의 기법이 데이터 자체에 내재된 어떤 성질을 충실히 찾아내는 것보다, 주어진 가정이 잘 들어맞는지 확인해 보는 통계적 기법에 의존했기 때문이다.

그러나, 딥러닝<sup>1)</sup> 기술이 보편화되면서 이러한 점은 크게 달라졌다. 딥러닝은 사람의 개입이 최소화되는 방식으로 작동되며, 과거 전통적 기법이 비교적 단순한 형태로 모델링되었던 것에 비해 이제는 수십억 개의 매개변수를 사용해 모델링 할 수 있게 되었다.

딥러닝의 이러한 비약적인 발전은 크게 두 가지 변화를 가져왔는데, 첫째는 이제 데이터의 양에 비례하여 충실히 그 성능과 효과를 얻을 수 있다는 점이고, 다른 하나는 비정형화<sup>2)</sup>된 소위 빅데이터를 효과적으로 다룰 수 있는 도구가 생겼다는 점이다.

---

1) 딥러닝은 제6장-2절에서 자세히 설명한다.

2) 정형화된 데이터와 비정형 데이터의 구분과 특성은 제2장-1절에서 자세히 설명한다.

## 2 빅데이터와 인공지능의 관계

인공지능은 데이터에서 유의미한 정보를 추출해내는 기법이다. 따라서 인공지능 기법은 데이터 없이는 불가능하다. 따라서 인공지능은 데이터가 많건 적건 반드시 데이터를 필요로 하고 데이터 없이는 불가능한 기술이다.

그렇지만, 모든 인공지능 기법이 반드시 방대한 데이터를 필요로 하는 것은 아니다. 인공지능 기법은 매우 다양하며, 그중 빅데이터의 주류를 차지하는 영상, 음성, 이미지 등<sup>3)</sup>을 다루기에 적합한 기술은 딥러닝 정도로 국한된다.

한편, 빅데이터를 다루기 위해서는 다수의 기술이 필요하다. 우선 그 방대한 크기로 인해 이를 효율적으로 저장하는 기술과 함께 효율적으로 처리하는 기술이 모두 필요하다. 한편으로는 저장된 빅데이터로부터 의미 있는 정보를 추출하기 위한 기술이 필요한데, 바로 이 지점이 인공지능 기법 중 특히 딥러닝이 필요한 지점이다. 빅데이터는 일반적인 데이터와는 다른 여러 특성을 가지고 있는데 이러한 특성을 다루기에는 딥러닝 기술이 상당히 유효하다. 이 때문에 최근에는 빅데이터를 다룬다고 할 때는 거의 딥러닝 기술이 사용된다는 것을 암시하는 경우가 많다.

요약하자면, 인공지능은 반드시 데이터를 필요로 하지만, 기법에 따라 반드시 빅데이터일 필요는 없으며, 빅데이터는 여러 가지 기법으로 다룰 수 있지만 최근에 각광받는 딥러닝 기법은 이미지나 영상, 음성 등을 다루기에 좋으므로 여러 관점에서 매우 적합하다. 결국 빅데이터와 인공지능 딥러닝은 거의 같이 언급되고 있다.

3) 이러한 데이터를 비정형 데이터라고 부른다. 정형과 비정형 데이터는 다음 절에서 자세히 살펴본다.

### 3 빅데이터는 얼마나 큰가?

데이터를 세는 가장 작은 단위는 비트(bit)이고, 8비트는 1바이트라는 것 정도는 이제 보편적인 상식이다. 그 다음 단위를 살펴보면, 1,024바이트<sup>4)</sup>가 모여 1킬로바이트가 되고, 10<sup>3</sup>킬로바이트는 다시 1메가, 10<sup>6</sup>바이트가 된다. 이 정도 단위도 많이들 들어 보았을 것이다. 그러나 빅데이터에 사용되는 단위는 기가 정도로는 어림도 없다. 다음 표를 보자.

〈표 1-1〉 데이터 명칭과 크기

명칭	크기(바이트)
기가(Giga)	1024메가
테라(Tera)	1024기가
페타(Peta)	1024테라
엑사(Exa)	1024페타
제타(Zetta)	1024엑사
요타(Yotta)	1024제타
브론토(Bronto)	1024요타

〈표 1-1〉은 데이터의 크기에 관한 명칭을 보여주고 있다. 빅데이터의 크기를 나타내는 용어들은 〈표 1-1〉에서 보는 것처럼 시간이 갈수록 계속 더 생겨나고 있다.

페이스북(Facebook)은 2021년 기점으로 하루 4페타바이트 이상의 데이터를 생산한다. 4페타바이트는 100기가바이트 크기의 하드디스크가 무려 41,943개나 필요한 엄청난 용량이다.

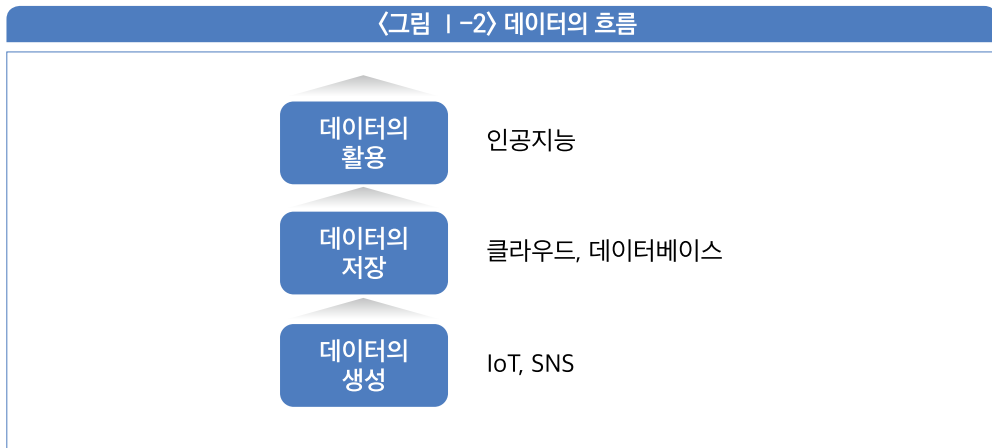
유튜브(YouTube) 역시 유사한 정도의 새로운 데이터를 생성하는 등 전 세계에서 쏟아지고 있는 데이터의 규모는 상상을 초월한다. 이처럼 빅데이터란 기존의 단위를 훌쩍 뛰어넘는 엄청난 규모로 다양한 곳에서 생산되고 있는 것이다.

4) 1킬로바이트는 2<sup>10</sup>바이트이며, 따라서 1,000바이트가 아니라 1,024바이트이다.



## 4 데이터의 흐름

이 책은 인공지능, 클라우드<sup>5)</sup>, 사물 인터넷(IoT)<sup>6)</sup>을 다루고 있다. 이 세 가지 기술은 데이터의 생성에서 최종 가공에 이르는 경로에서 서로 유기적으로 연계되어 있다. 다음 그림을 보자.



〈그림 1-2〉은 데이터의 수집부터 가공 그리고 응용에 이르기까지의 흐름을 같이 보여주고 있다. 그림에서 보는 것처럼 인공지능이나 빅데이터 기술은 서로 완전히 독립된 상태로 구동되는 것이 아니라 연계된 단계에 따라 각자 밀접하게 연계되어 있다.

〈그림 1-2〉을 보면 먼저 사물 인터넷의 각종 센서 또는 SNS 프로그램을 통해 다양한 데이터가 지속적으로 생성되며, 이렇게 생성된 대량의 데이터가 저장소로 전송되는데, 이에 대한 유연한 대응을 위해 주로 클라우드를 이용해 저장한다. 이렇게 클라우드에 저장된 데이터는 그 후속단계에서 인공지능 등의 다양한 기술을 통해 이를 보다 다양한 응용분야에 맞도록 가공하는 단계가 이어진다.

5) 클라우드는 제2장에서 자세히 다룬다.

6) 사물 인터넷은 제3장에서 자세히 다룬다.

대부분의 회사에서 통상 주로 사용하고 있는 데이터처리 시스템은 데이터베이스 시스템이다. 데이터베이스는 데이터의 저장은 물론 주어진 조건에 부합하는 데이터를 효율적으로 찾아내는 것은 물론 데이터가 허용된 범위를 벗어났는지 여부를 검사하는 무결성 점검 등의 다양한 기능을 제공한다.

이러한 데이터베이스 중 가장 보편적으로 사용되고 있는 것은 바로 ‘관계형(Relational)’ 데이터베이스이다. 관계형 데이터베이스는 잘 짜여진 형태로 구성된 테이블 방식으로 데이터를 저장하고, 각 테이블에 저장된 데이터 사이의 유기적인 ‘관계’를 설정하고 이용할 수 있는 방식으로 데이터를 저장한다. 독자 여러분들이 일상적으로 사용하고 있는 엑셀 파일을 생각해 보면 쉽게 이해 될 수 있을 것이다. 엑셀 파일은 일정한 틀을 가진 규격화된 열과 행에 숫자나 기타 데이터를 입력한다. 이때 대개 첫 번째 행에는 각 열이 어떤 특성의 데이터를 저장하고 있는지 그 속성을 알려주는 속성 이름을 지정한 값을 저장해 둔다. 또한 각 행의 첫 열에는 각 데이터를 고유하게 특정할 수 있는 어떤 식별자를 두거나 그러한 식별자가 없을 때는 단순히 일련번호를 부여하기도 한다. 다음 그림을 보자.

〈그림 1-3〉 정형화된 데이터

이름	나이	성별
홍길동	33	1
김갑수	35	1
김영희	22	2

〈그림 1-3〉는 우리에게 매우 익숙한 엑셀 파일 데이터의 일부를 보여준다. 그림에서 보는 것처럼 이 데이터는 각 열에는 특정 속성이 기록되고, 행에는 각각 속성의 실제 데이터 값이 들어 있다. 이처럼 형식을 잘 갖추어 구성된 데이터를 흔히 구조화(Structured) 또는 정형화 데이터라고 한다. 이러한 잘 구성된 데이터는 주로 파일단위로 저장되어 있거나 데이터베이스에 테이블 형태로 잘 정리되어 보관된다.

반면 이러한 정형화된 형식을 갖추지 못한 일정한 형식이 없는 임의형식의 데이터를 비구조화 혹은 비정형 데이터라고 부른다.

대부분의 빅데이터는 비정형 혹은 비구조화의 속성을 갖고 있다. 다음 그림을 보자.

〈그림 1-4〉 전형적인 비구조화 데이터인 이미지



〈그림 1-4〉는 전형적인 비구조화 데이터의 일종인 사진 이미지를 보여주고 있다. 우리 주변에는 이러한 비정형 데이터를 아주 쉽게 발견할 수 있다. 페이스북에 올린 게시 글들, 유튜브에서 볼 수 있는 각종 동영상 자료들, 음악 앱에서 들리는 노래 소리 등은 모두 어떤 특정 틀을 가지고 있지 않은 비정형 데이터 들이다.

비정형 데이터는 이처럼 어떠한 일정 형식이 없이 각각 저마다의 독특한 특수 형태로 저장되며, 대체로 그 크기 또한 매우 크며 동시에 생성 빈도 역시 매우 빈번하다는 특성을 가지고 있다. 그럼, 이제 빅데이터만이 가지는 주요 특징을 한번 정리해 보도록 하자.

앞서 빅데이터는 대개 비정형 구조를 가진다는 사실을 살펴보았다. 빅데이터는 또한 3가지 특징을 가지는데 이는 모두 영문 V로 시작하는 단어로 기술할 수 있으며, 이 때문에 3V 속성이라고 부르기도 한다. 이제 3V 속성이란 무엇인지 하나씩 살펴보도록 하자.

### 1-1 속도(Velocity)

빅데이터의 첫 번째 특징은 속도(Velocity)이다. 여기서 속도라는 용어가 사용된 이유는 빅데이터가 빠르게 생성되는 모습을 묘사하기 위한 것이다. 앞서 잠시 언급한 것처럼, 비정형 빅데이터는 각종 센서나 인공위성, 그 밖의 다양한 도구로부터 생산된 방대한 데이터 혹은 여러 SNS를 통해 양산된 데이터 등이 축적되므로, 빅데이터가 생성되는 속도는 과거에 사람이 개입되어 시간과 비용을 들여 전처리를 함으로써 생성되었던 일반적인 정형화된 데이터들과는 비교할 수 없을 만큼의 빠른 속도로 생성되며, 이를 특징짓는 첫 번째 단어가 바로 속도를 나타내는 V이다.

### 1-2 볼륨(Volume)

빅데이터의 두 번째 성질인 볼륨은 빅데이터의 ‘빅(Big)’이라는 단어 그 자체를 의미한다. 앞서 여러 번 언급한 것처럼, 각종 센서나, 카메라 등에서 포착되고 생성되는 데이터의 용량은 상당히 크다. 대체로 1시간 분량의 동영상을 생성하고 저장하기 위해서는 그 해상도에 따라 다르지만 일반적으로 수백 메가바이트에서 부터 많게는 수십 기가바이트가 필요하다.

국내 공공기관에 설치된 CCTV는 2019년 기준으로 약 1,148,770대로 추산된다.<sup>7)</sup> 이 CCTV가 24시간 생성하는 데이터를 저장한다고 생각해 보면 어느 정도 분량의 데이터가 생성되고 있는지 어느 정도 감이 올 것이다. 또, 민간기관이나 가정에서 설치한 CCTV, 그리고 해외의 CCTV, 그리고 지구를 뺄곡이 둘러싸고 있는 인공위성으로까지 그 범위를

7) [https://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx\\_cd=2855](https://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx_cd=2855)

넓히고, 각 개인들이 유튜브 등에 업로드하는 수 많은 영상 등을 생각해보면 빅데이터의 볼륨이 의미하는 바를 보다 쉽게 이해할 수 있을 것이다.

### 1-3 다양성(Variety)

빅데이터의 마지막 특성은 그 다양성에 있다. 대개 테이블 형식으로 잘 정리된 형태가 일정한 소위 정형화 데이터와 달리 빅데이터는 동영상, 이미지, 음성 등 일정하지 않은 다양한 형태로 생성된다는 특징을 가지고 있다.

페이스북에 올리는 다양한 글들, 유튜브에 올리는 영상들, 인공위성이 찍어 내는 영상들, 우리가 말하는 음성들, 이 모든 데이터들은 일정한 형식을 가진 것이 아니라 제 각각 저마다의 형식으로 생성되고 저장된다. 이처럼 빅데이터는 무한한 형태로 생성될 수 있으며, 각각의 응용에 따라 데이터를 처리하고 맞추어 줘야 한다는 특징을 가지고 있다.

한편, 최근에는 빅데이터의 기본적인 이 세 가지 속성에 더해 ‘바람직한 속성’ 두 가지를 더해서 모두 5V로 표현하기도 한다. 추가적인 두 가지 V는 각각 정확성과 가치에 대한 것인데, 하나씩 살펴보도록 하자.

### 1-4 정확성(Veracity)

정확성은 생성되는 데이터의 신뢰와 관련된 지표이다. 빅데이터의 대부분이 비정형화 되어 있고 서로 상관이 없는 데이터이므로 이를 적절히 걸러내어 오직 비즈니스에 상관 있는 데이터만을 적절히 얻어내는 방법이 빅데이터의 또 다른 주요한 추가적인 속성으로 대두되고 있는 것이다. 너무 많은 데이터는 때로는 너무 적은 데이터보다 더 적은 정보를 줄 수 있다. 따라서 적절한 처리를 통해 처리해야 할 데이터 범위를 축소할 수 있다면 효율의 향상은 물론 보다 질 좋고 적절한 정보를 얻을 수 있는 확률은 비례적으로 향상될 것이다.

## 1-5 가치(Value)

또 다른 빅데이터의 추가적인 성질은 그 가치이다. 즉, 단순히 더 많은 데이터를 축적한다고 해서 반드시 비즈니스에 도움이 되는 것은 아닐 수도 있다는 점은 자명할 것이다. 이 때문에 양이 아닌 질적인 가치를 추구하여 적절히 이를 통제하고 분석할 수 있는 방법론이 필요하며, 따라서 보다 가치 있는 데이터를 걸러낼 수 있는 기능이 빅데이터 처리에 있어 무엇보다 중요한 속성이 된다.

## 2 금융과 빅데이터

금융에서는 어떠한 종류의 빅데이터가 존재할까? 금융 분야에서는 본질적으로 다양한 종류의 빅데이터가 생성된다. 수 많은 금융거래 고객의 데이터는 기본이고 전 세계주식시장의 거래 데이터, 자금 이체, 신용 카드 결제 데이터가 발생한다. 비자카드의 경우 초당 4,000건 이상의 신용 카드 결제 처리가 가능하며, 우리나라 금융결제원은 연간 100억 건 이상의 자금 거래를 처리한다. 또한, 직접적인 금융 데이터에 더해 각국 중앙은행의 발표문, 기업들의 재무제표, 각종 언론들의 언급, 사람들의 의견 등 간접적인 지표를 구성하는 다양한 데이터 또한 존재한다. 최근에는 위성 사진으로 찍은 특정 지역의 모습 등도 경제의 방향성을 예측하는 데 활용되는 등 점점 더 다양하고 거대한 데이터 들이 금융시장 예측에 사용되고 있다.

### 2-1 플래시 크래시

2010년 5월 6일 오후 2시 45분<sup>8)</sup> 미국 주식 시장에서 다우존스 산업 평균지수가 순식간에 무려 10% 가까이 폭락했다. 그런데 더 놀라운 것은 지수가 불과 몇 분만에 다시 대부분 손실을 회복할 정도로 반등했다는 사실이다. 미국 감독 당국은 이 예상치 않은 폭락을

8) 미국 동부 시각 기준, 서머타임 적용

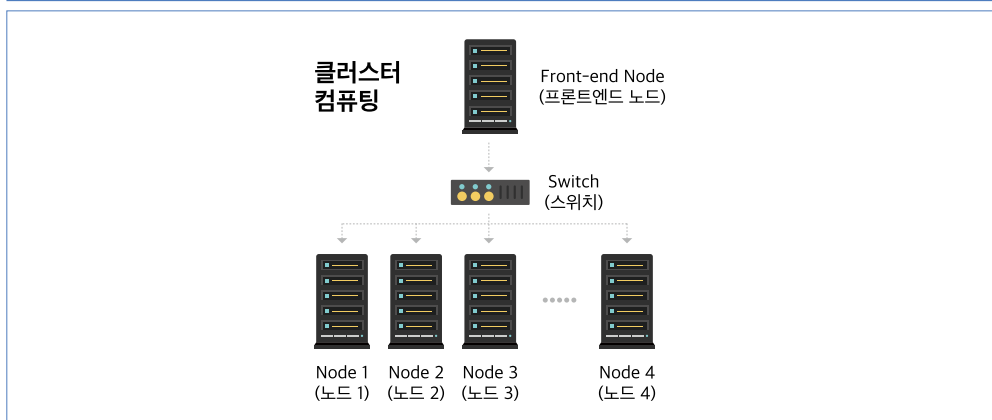
조사하는 데 무려 5개월을 허비했다. 미국 감독 당국은 느린 조사에 대한 의회에서의 질타에 대해 거래 자료가 워낙 방대해서 조사하는 데 막대한 시간이 소요될 수밖에 없었다고 강변했다. 당시 미국 감독 당국이 조사해야 했던 데이터는 약 20테라 바이트 정도인 것으로 알려져 있다.<sup>9)</sup>

## 2-2 하둡(Hadoop) 과 HPC(High Performance Computer)

빅데이터를 처리하기 위한 하드웨어로 주로 하둡과 HPC가 이용된다. 이 두 아키텍처의 공통점은 슈퍼 컴퓨터가 아닌 일반 PC를 클러스터로 묶어 슈퍼 컴퓨터에 맞먹을 수 있는 성능을 내도록 하는 데 초점이 맞춰져 있다는 것이다.

빅데이터를 다루기 위해서는 슈퍼 컴퓨터에 맞먹는 계산 자원과 함께 그에 걸맞는 거대한 저장공간이 필요하다. 그러나 일반 기업에서 빅데이터 처리를 위해 슈퍼 컴퓨터를 반드시 구매해야만 한다면 어느 정도 이상의 수준과 규모를 갖춘 대기업이라 하더라도 선뜻 엄두를 내지 못할 수도 있을 것이다. 세계 5대 슈퍼 컴퓨터는 대당 1,000억원을 훌쩍 넘는 고가이며, 일반적인 슈퍼컴퓨터라 하더라도 최소 몇 백억 정도의 비용이 소요되기 때문이다.

〈그림 1-5〉 일반적인 클러스터



9) 실전 금융머신러닝 완벽분석, 마르코스 로페즈 드 페라도저, 이병욱 역, 에이콘 출판사. P.476

〈그림 1-5〉는 일반적인 클러스터(Cluster)의 구성 모습을 보여준다. 클러스터란 동일한 목적을 위해 군집을 이룬 일단의 컴퓨터 그룹을 의미한다. 대개 동일한 하나의 과제를 서로 분산해서 처리하므로 클라우드와는 구분된다. 빅데이터 처리를 위해서는 하둡이나 HPC 방식의 클러스터를 형성한 후 데이터를 분산하여 효율적으로 처리한 다음 역시 이를 종합해서 처리한다. 그렇지만 파일 저장 방식은 둘이 서로 조금 다른데, 일반적으로 하둡은 파일을 분산하여 저장하는 것이 그 핵심이지만 HPC에서의 파일은 하둡에서보다 훨씬 더 크며 대개 중앙화되어 한 곳에 저장된다. 또 HPC에서는 하둡에 비해 보다 빠른 통신 장비가 요구된다.

이때 클러스터에 사용되는 컴퓨터는 사내의 유휴 컴퓨터를 사용하거나 일반형 컴퓨터로 사용해도 충분할 수 있도록 구성된다. 때문에 빅데이터를 처리하기 위한 플랫폼 구성에 그렇게 많은 비용이 소모되지 않는다. 빅데이터를 저장하고 처리하며 운영하는 방법은 제 3절에서 자세히 살펴보기로 하자.



빅데이터를 다루는 데 있어 일반적인 데이터와 가장 차이나는 부분은 이름에서 알 수 있는 것처럼 단연 그 크기이다. 우리가 사용하는 대부분의 일상적 데이터는 파일단위로 데이터를 저장하고 읽는 데 아무런 문제가 없다. 예를 들어, 우리의 일상 업무에서는 엑셀 파일이나, 워드 파일 등을 통해 작업을 수행하고 그 단위로 저장한다.

2021년 최신 버전의 마이크로소프트 엑셀 파일이 저장할 수 있는 최대 데이터는 1,048,576개의 행과 16,384의 열이다. 이 범위를 벗어난 데이터는 하나의 엑셀 파일에 저장하는 것이 불가능하다.

빅데이터의 크기는 통상적인 범위를 벗어나고 이는 별도의 처리 방법이 필요하다. 이 절에서는 빅데이터를 저장하고 처리하는 일반적인 방법에 대해 간략하게 살펴보도록 하자.

**TIP**

이 절에서 다루는 내용은 다소 기술적인 내용이며 전문용어가 다수 나타난다. 이 절의 내용이 다소 어려워 완전히 이해가 안 되어도 걱정할 필요는 없다. 키워드 위주로 어떤 단어가 사용되었는지 정도만 이해해도 큰 지장은 없다.

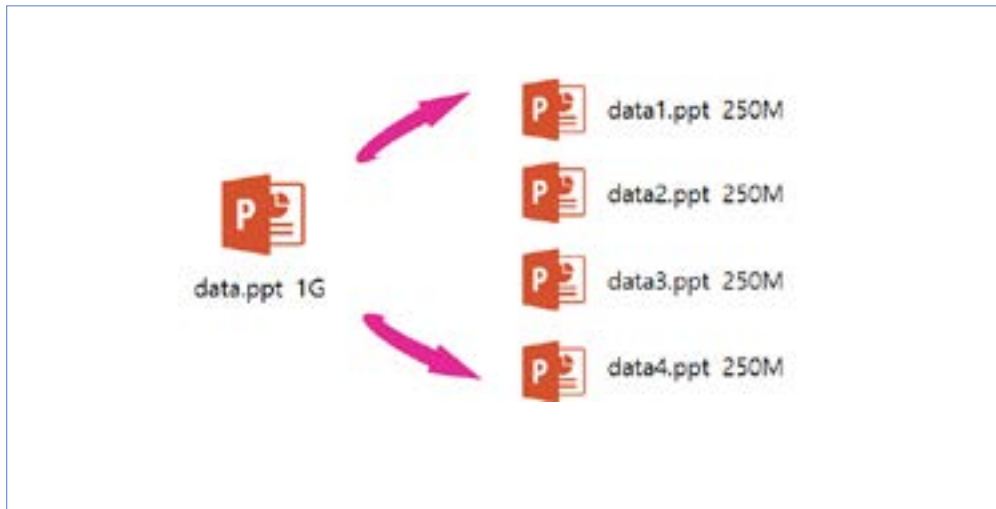
## 1 빅데이터 저장 기술

빅데이터를 저장하기 위해서는 일반적으로 사용하는 것과는 다른 방식이 필요하며, 다양한 연구들이 진행되고 있다. 여기서는 빅데이터 저장 방식의 특별한 점을 알아보는 관점에서 그중에서 HDF5라 불리는 데이터 형식을 소개하기로 한다.

### 1-1 분산파일 시스템

우리는 작성한 문서를 대개 하나의 파일에 저장한다. 워드나 엑셀 작업이 완료되면 해당 파일 이름을 고유하게 지정한 다음 각 파일에 독립적인 내용을 담고 적절한 폴더를 지정해 가지런히 저장한다. 그러나 간혹 파워포인트 등의 프리젠테이션 파일에 여러 동영상 등을 삽입하다 보면 그 용량이 매우 커져서, 수백 메가바이트, 혹은 수 기가바이트 단위로 까지 커질 때면, 이를 하나의 파일로 만든 다음에 이메일로 전송한다거나, 특정 사이트에 업로드하려 시도할 때 그 용량의 제한으로 인해 제대로 작업을 수행하지 못해 곤혹을 치른 경험도 있을 것이다. 이 경우, 우리는 대개 대용량 파일을 여러 개로 분리해 저장하기 마련이다.

〈그림 1-6〉 대형 파일의 저장(예시)



〈그림 1-6〉을 보자. 그림에는 1기가바이트나 되는 data.ppt 파일을 4개의 250M 바이트 파일로 나누어 저장하고 있는 모습을 보여준다. 이때 파일을 몇 개로 나눌지, 크기를 어떻게 할당할지, 나눈 파일의 이름을 어떻게 정할지는 모두 사람이 직접 개입되어 실행한다. 또 막상 파일을 나눈 다음에, 어느 내용이 어느 파일에 있었는지 기억해야 하는 등 관리가 매우 힘들다.

빅데이터의 분산 파일 시스템은 바로 앞서의 모든 과정을 자동화하는 것에 비유할 수 있다. 즉, 파일이 분산되어 저장되고 어느 파일에 어떤 내용이 있는지, 필요한 경우 이 파일을 합치고, 검색할 경우 어느 파일에 있는지 검색해서 데이터를 찾는 등의 모든 과정이 자동화되어 실행되는 시스템이 필요한 것이다. 그럼 빅데이터에서 이러한 분산 파일 시스템 중 하나인 HDF5에 대해서 간략히 살펴보기로 하자.

## 1-2 HDF(Hierarchical Data Format)

HDF는 대용량 데이터를 저장하기 위해 설계된 파일 형식으로서 계층적 구조를 가지고 있다. 이 형식의 개발은 “The HDF Group<sup>10)</sup>”에서 주도하고 있다. “The HDF Group”은 비영리 재단으로서 HDF를 사용할 수 있는 소프트웨어는 무료로 배포되고 있으며 여러 플랫폼과 컴퓨터 언어에서 이 형식을 지원하고 있다.

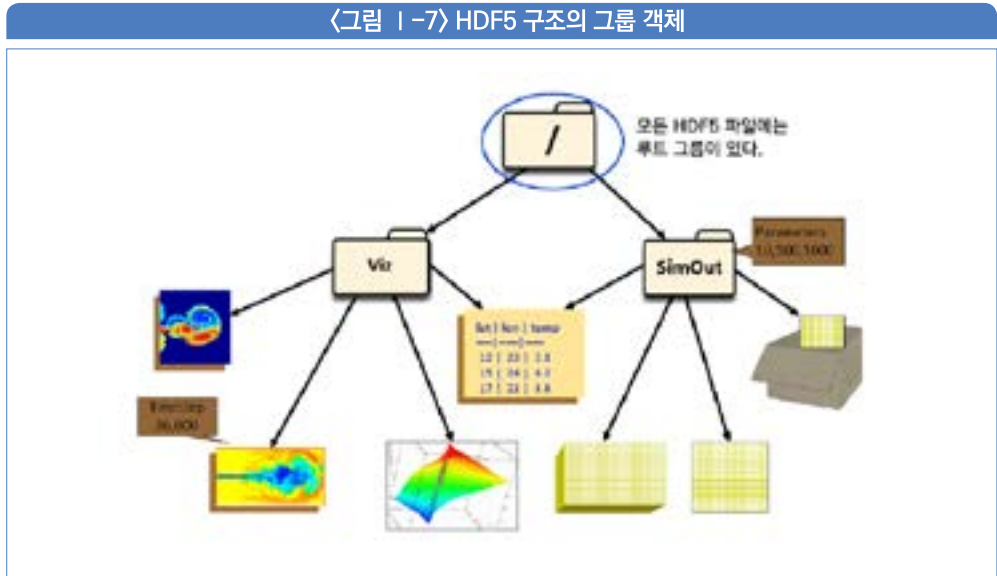
2021년 현재의 버전은 기존의 HDF4를 크게 개선한 HDF5 형식을 사용하고 있다. HDF5는 단순히 데이터 형식을 의미하는 것은 아니고, 데이터 모델, 라이브러리, 데이터를 저장하고 조작하기 위한 파일 형식을 모두 통칭하는 용어이다.

HDF5의 가장 큰 특징은 파일 속에 파일이 들어 있는 구조라는 것이다. 즉, 기존에 폴더와 파일로 구성된 일반적인 구조가 파일 하나에 다 들어와 있는 계층 구조를 상상하면 된다. HDF5의 데이터 모델에는 다양한 객체가 존재하는데 여기서는 가장 중요한 두 객체로서 그룹과 데이터셋을 살펴보기로 한다.

10) 회사에 대한 소개는 <https://www.hdfgroup.org/>를 참고하라.

## 가. HDF5 그룹 객체

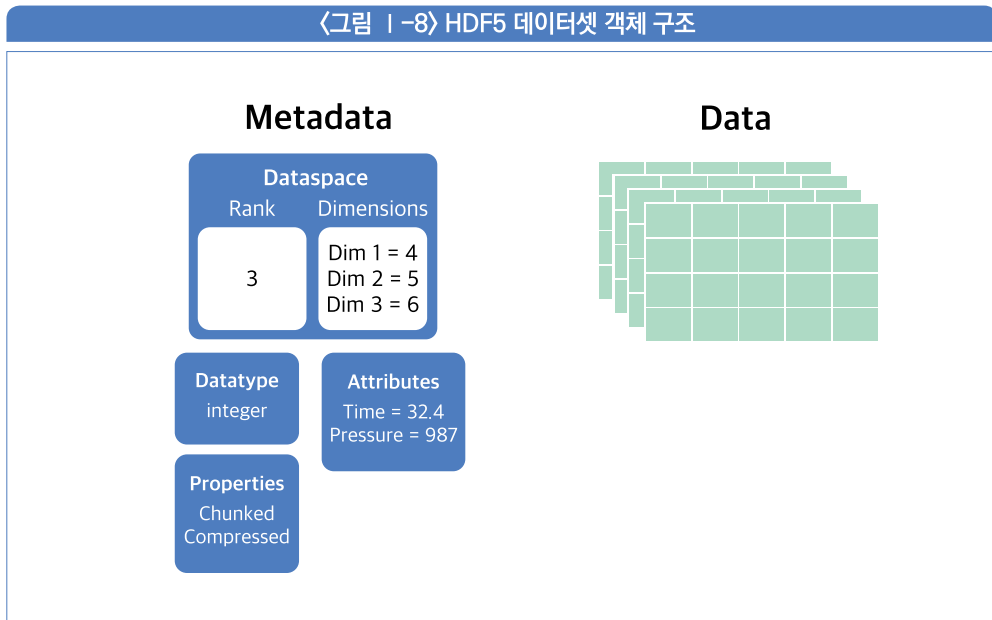
먼저 그룹 객체부터 살펴보기로 하자. 다음 그림을 보자.



출처: The HDF group 홈페이지

〈그림 1-7〉은 HDF5 구조의 그룹 객체를 보여준다. 그룹 객체는 데이터 객체를 유기적으로 구성하고 있다. 모든 HDF5 파일은 다른 그룹을 포함하거나 혹은 다른 파일의 객체로 연결될 수 있는 루트 그룹을 가지고 있다. 〈그림 1-7〉의 최상단에 보이는 폴더 그림이 바로 루트 그룹이다. 마치 일반적인 컴퓨터 디스크의 폴더 구조처럼 보이지만, 사실 모두 단일 파일 하나에 들어 있는 구조이다. 즉, 폴더로 이루어져 있는 복잡한 구조가 하나의 파일이라고 생각하면 된다. 다시 말해, 파일 내부에서 폴더와 같이 여러 분류 단계를 둘 수 있고 각각 분류 단계 내에 또 다른 데이터를 저장하여 계층 구조를 형성할 수 있는 것이다.

## 나. HDF5 데이터셋 객체

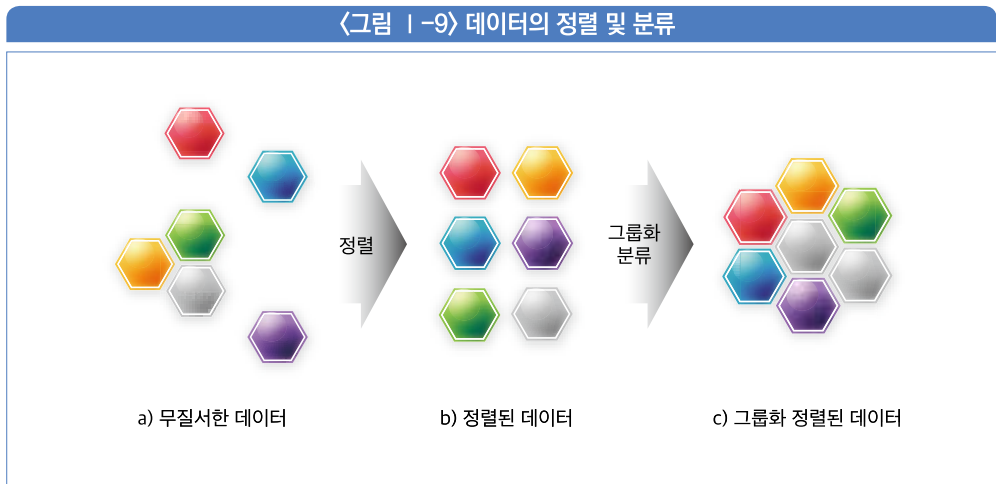


출처: HDF5 홈페이지

<그림 1-8>은 HDF5의 데이터셋 객체 구조를 보여주고 있다. 데이터셋은 원시 데이터 값을 저장하여 구성하고 있다. 그림의 오른쪽은 실제 데이터를 포함하고 있고 그림의 왼쪽은 이 데이터의 여러 특성을 설명해주는 메타데이터로 구성된다. 자세한 설명은 생략하지만 이 형식에서는 데이터와 그 데이터가 가진 성질을 기술해 메타데이터라는 두 부분으로 구성된다는 점을 기억하면 된다.

## 2 빅데이터 처리 기술

앞서 빅데이터를 저장하고 추출하는 분산파일 시스템에 대해 간략히 살펴보았다. 실제로는 빅데이터는 수집된 그대로 저장되고 않고, 효율적인 처리를 위해 전처리를 거쳐 정렬과 분류 작업을 거친다. 다음 그림을 보자.



<그림 1-9>의 a)는 수집된 그대로의 무질서한 상태를 보여준다. 엄청난 양의 무질서한 데이터 그대로 의미 있는 정보를 추출하는 것은 거의 불가능하다. 따라서, 이러한 데이터를 우선 정해진 기준에 따라 일목요연하게 정리할 필요가 있다. 그림의 b)는 어떤 기준에 따라 잘 정렬된 데이터를 보여준다. 또한 이 방대한 데이터는 중복은 물론, 유사한 것들이 많이 모여있을 것이므로 이를 또 다시 특정 기준에 따라 묶어둘 수 있다면 보관과 추출에 있어 매우 유용할 것이다. 그림의 c)는 이러한 과정을 거쳐서 그룹으로 정리된 데이터를 보여준다. 일반적으로 정리 단계를 거치게 되면 다수의 중복 등이 제거되므로 데이터는 상당히 축소된 상태가 된다.

빅데이터에서 이러한 정렬 및 정리를 위해서는 몇 가지 프레임워크가 있는데, 그중 널리 쓰이는 것 중 하나는 맵리듀스(MapReduce)이다.

맵리듀스는 구글과 같은 검색 엔진 업체들이 전 세계에서 수집한 방대한 데이터를 효율적으로 다루기 위한 연구에서 출발하여 발전 및 생성된 기법이다. 구글 검색엔진이 수집하여 저장하고 있는 전 세계의 방대한 데이터들을 상상해 보면, 이 거대한 빅데이터를 효과적으로 저장하고 추출하는 처리법이 필수적이라는 것이 쉽게 이해될 것이다.

맵리듀스의 핵심은 병렬로 처리 가능한 여러 문제들을 대량의 컴퓨터로 이루어진 클러스터에서 효과적으로 초대형 데이터셋을 처리할 수 있는 방법론에 대한 것이다.

맵리듀스는 맵(Map)과 리듀스(Reduce)를 합친 합성어로서 앞의 <그림 1-9>에 개념적으로 비유하자면 맵은 정렬을 위해 원시 데이터를 기준과 매칭시키는 과정이고, 리듀스(Reduce)는 그 이름에서 암시하듯 일련의 범주화 과정을 거쳐 전체 데이터를 상당 부분 줄여서 저장한다.

#### MEMO

빅데이터 처리에 보편적으로 사용되고 있는 대표적인 시스템은 하둡(Hadoop)이다. 하둡은 구글의 분산 파일시스템(GFS)과 맵리듀스(MapReduce)를 통합 구현한 것이며, 구글 분산 파일시스템은 하둡에서 HDFS(Hadoop Distributed File System)으로 구현돼 있다.

하둡을 가장 활발하게 사용하고 있는 기업은 페이스북이다. 페이스북이 하둡을 도입한 것은 2004년이며, 현재 4,000여 개의 기계를 클러스터로 구성한 세계 최대 규모의 하둡을 운영 중에 있다. 페이스북은 로그의 처리와 추천시스템, 데이터 웨어 하우스 등의 광범위한 용도로 하둡을 사용하고 있다. 이 밖에도 야후, 아마존 등은 검색엔진에 하둡을 이용하고 있고, 야후는 로그처리에도 하둡을 이용하고 있다. 뉴욕 타임스 등의 언론사도 비디오와 이미지 분석에 하둡 시스템을 도입하고 있다.

한 가지 기억할 것은 빅데이터를 다루기 위한 여러 도구를 일반 데이터에 적용하면 오히려 성능이 저하될 수 있다는 점이다. 일반적인 용량의 정형화 데이터를 처리하는 데는 빅데이터를 위한 아키텍처가 아니라 관계형 데이터베이스가 월등히 더 성능이 좋다.

### 3 빅데이터 비즈니스 사례

빅데이터의 비즈니스 사례에서는 대개 인공지능 기법이 이용된다. 이는 앞서 설명한 빅데이터와 인공지능의 관계로부터 쉽게 그 이유를 유추할 수 있을 것이다. 따라서 여러 사례는 제4절의 인공지능의 이해에서 다루어질 것이다. 여기서는 그럼에도 불구하고 대표적으로 알아둘 세 가지만 간략히 살펴본다. 하나 기억할 것은 여기에서 소개되는 세 가지 역시 대부분 인공지능 기법을 이용해 수행된다는 사실이다.

#### 3-1 군집화

군집화(Clustering)는 데이터의 각 포인트를 검사하고 이들을 정해진 기준에 따라 그룹화하는 절차로서 각 데이터 포인트는 클러스터(Cluster)<sup>11)</sup> 단위로 자동 분리된다. 군집화는 머신 러닝 방법으로 수행되며 정답이 없는 데이터로 수행되므로 앞서 배운 것처럼 비지도 학습<sup>12)</sup>의 일종이다. 군집화에는 대표적으로 계층적 방식과 할당 방식이 있다.

##### 가. 계층적 군집화

계층적 군집화 모델에서는 모든 데이터 포인트가 하나의 클러스터가 되어 시작한다. 각 클러스터는 사전에 정한 기준에 의해 주위의 클러스터와 합쳐져 새로운 클러스터를 형성하며, 군집화를 반복한다. 이 때문에 계층적 군집화는 병합적 군집화라고도 부른다. 군집화는 사전에 정한 개수의 클러스터가 생성되었을 경우 중단한다.

##### 나. 할당적 군집화

할당적 군집화 모델에서는 각 점을 특정 척도에 의해 가장 적합한 그룹에 배치시킨다. 일반적인 모델은 유클리드 거리 개념이나 비유사도(dissimilarity)와 같은 비용함수(cost function)를 사용하는 것이며, k-평균 알고리즘이 가장 많이 사용된다. 비용함수란 목적함수(objective function)라고도 불리는데, 어떠한 알고리즘이 목표를 삼는 방향을

11) 여기의 클러스터는 단순히 '군집'의 영어명칭일 뿐 앞서 설명한 하둠 등의 클러스터를 의미하는 것은 아니다.

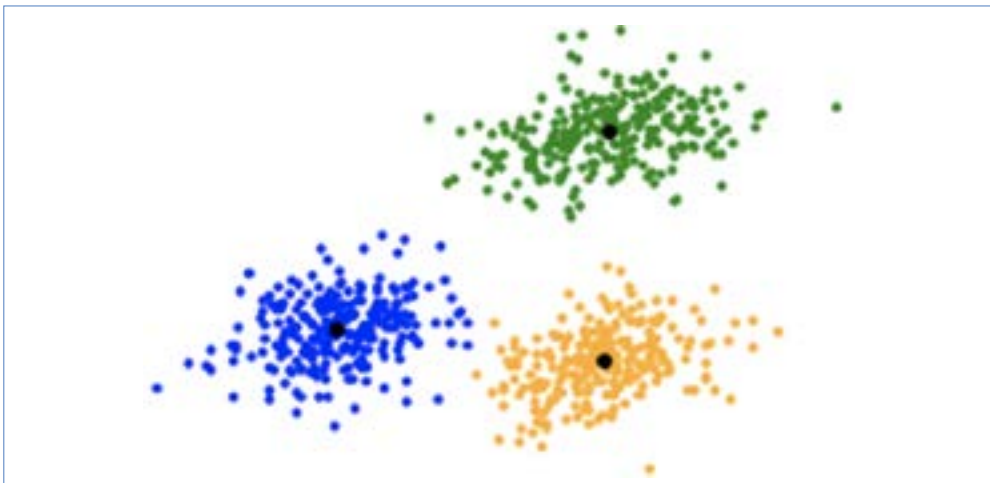
12) 비지도 학습은 제5절 2에서 자세히 설명한다.



의미하며, 이 비용함수의 값을 최소화 또는 최대화하는 것을 목표로 한다. 즉, 비용함수 값이 최소화(또는 최대화) 되면 원하는 최적의 해를 찾을 수 있다고 믿는 것이다.

k-평균 알고리즘은  $n$ 개의 입력 데이터를  $k$ 개( $k < n$ )의 그룹으로 나누는데, 각 그룹 내의 데이터들은 매우 유사(즉, 가까운 거리)해야 하며, 하나의 그룹에 속한 데이터 포인트와 다른 그룹에 속한 데이터 포인트는 최대한 유사하지 않게(즉, 먼 거리) 될 수 있는 그룹화를 찾는다. 일반적으로 하나의 그룹의 중심점(centroid)을 정하고, 이 중심점과 모든 데이터 포인트와의 거리를 척도로 사용한다.

〈그림 1-10〉 k-평균 군집화의 예시



〈그림 1-10〉은 k-평균 군집화의 예시를 보여준다. 그림에서는 3개로 군집화되는 모습을 보여주고 있다. 각각 클러스터의 중심부에 보이는 까만 점이 바로 중심점(centroid)이다. 최초의 중심점은 임의의 위치(대개 특정 데이터 포인트 하나를 임의로 뽑아 그 점을 중심점으로 삼음)에서 출발하며, 각 데이터 포인트는 자신에게서 가장 가까운 중심점에 할당된다.

모든 할당이 완료되면, 각 클러스터의 중심점이 이동하게 되는데, 자신에게 할당된 모든

점의 평균위치가 새로운 중심점이 된다. 중심점이 이동했으므로, 다시 전체 데이터 포인트에 대해 각 클러스터의 새로운 중심점과 자신과의 거리를 계산해 더 가까운 곳으로 재배치된다. 재배치 후 다시 앞서의 방법으로 중심점이 재계산된다. 이는 더 이상 새로운 배치가 일어나지 않거나 중심점이 이동하지 않으면 멈춘다. 할당적 군집화는 분할적 군집화로도 불린다.

### 3-2 추천시스템

추천시스템은 군집화 알고리즘에 기반하고 있다. 추천은 기본적으로 유유상종이라는 철학에 바탕을 두고 있다. 즉, 유사한 유형의 사람들은 그 행동에 있어서도 서로 공통점이 많을 것이라 판단하는 것이다. 동영상 사이트나, 쇼핑몰, 서점 등에서 자신에게 맞춤형 추천 상품을 보여줄 때의 숨은 인공지능 알고리즘은 바로 ‘나를 비슷한 유형의 집단 중 하나로 분류해 소속시킨 다음, 그 유형에서 자주 일어나는 행동 양식을 나에게도 추천하는 방식으로 이루어진다. 예를 들어 로맨스 장르의 영화를 주로 보는 사람을 A 그룹으로 분류하고 호러 영화를 주로 보는 사람들을 B 그룹으로 분류한다면, A 그룹에서 자주 시청하는 영화의 종류와 B 그룹에서 선호하는 영화의 종류는 다를 수 있고, 만약 어떤 시청자가 A 그룹으로 분류된다면 그는 기존에 A 그룹이 즐겨보던 영화 종류를 추천받을 가능성이 점진적으로 높아지게 되는 것이다.

이러한 추천시스템은 많은 경우 구매 욕구를 강하게 자극하여 매출 향상에 크게 도움이 되므로, 여러 기업들은 보다 정교한 추천 알고리즘을 찾기 위해 많은 투자를 아끼지 않고 있다.

한편 추천시스템에서는 특정 상품을 구매한 사람이 연관된 다른 상품을 구매하는 패턴을 정교하게 분석하는 방법도 많이 사용되는데, 고객이 거래한 구매데이터로부터 특정 구매 패턴을 추출해내는 방식을 보통 장바구니 분석(Market basket analysis)라고 부른다. 이 장바구니를 분석하면 그 결과는 연관규칙(association rule)의 집합으로 이루어지는데, 예를 들면 다음과 같다.

### {삼겹살, 상추, 쌈장} → {소주}

즉, 이 규칙은 삼겹살과 상추, 쌈장을 구매한 경우 대개 소주를 구매할 가능성이 있다는 것을 보여준다. 따라서 실제 발생한 모든 구매 패턴으로부터 위와 같은 규칙을 뽑아낼 수만 있다면 상당히 높은 확률로 고객들이 구매하는 물품끼리의 상호 연관성을 찾아낼 수 있는데, 이때 가장 어려운 점이 바로 검색해야 할 데이터의 용량이 너무 크다는 점이다. 예컨대  $k$ 개의 아이템에 대해 모두 조사한다면 이론적으로  $2^k$ 개의 가능한 아이템 집합이 존재한다. 즉 쉽게 빅데이터가 되어 버린다는 것이다. 이 때문에 대개 검색해야 할 대상을 줄여주는 아프리오리(Apriori) 알고리즘을 같이 사용한다. 아프리오리 알고리즘에 대한 자세한 설명은 여기서는 생략하지만 대체로 빈번하게 등장하는 아이템이 아닌 경우 검색대상에서 제외할 수 있다는 직관에서 비롯해 연관성이 낮은 데이터를 삭제함으로써 검색해야 할 대상을 대폭 줄여주는 알고리즘 정도로 생각하면 된다. 이 밖에도 지지도(Support)와 신뢰도(Confidence) 등의 통계적 척도를 동원해 규칙 검색의 효율성을 높이는데, 지지도 역시 전체 거래 데이터베이스에서 특정 아이템이 등장하는 빈도에 기반해 측정한다.

### 3-3 사회관계망 그래프 마이닝

그래프 마이닝(Graph Mining)은 데이터 마이닝의 그래프 버전으로 생각하면 된다. 그래프는 꼭짓점과 선분으로 이루어진 형태를 사용해 특정 객체들 사이의 관계를 시각적으로 표현하는데 사용되는 방식이다.

실생활의 여러 현상을 그래프로 모델링하고 이를 시각화하면 여러 가지 의미있는 정보를 추출할 수 있다. 빅데이터에서 가장 간단하고도 보편적으로 사용되는 그래프 마이닝 중 하나는 단어 구름(Word Cloud) 그래프이다.



## 1 인공지능의 역사

### 1-1 마빈 민스키와 프랭크 로젠블랫

인공지능의 발전 역사를 보면 현대 인공지능 기술의 태동기로 볼 수 있는 1950년대에 있어 두 명의 이름을 거론하지 않을 수 없으니 바로 마빈 리 민스키(Marvin Lee Minsky)와 프랭크 로젠블랫(Frank Rosenblatt)이다.

1절에서 잠시 살펴본 것처럼 인공지능이라는 말은 1956년 다트머스 대학에서 열린 컨퍼런스에서 처음으로 사용되었다. 이 컨퍼런스에는 마빈 민스키라는 미국의 과학자도 참석했으며, 그는 1958년부터 MIT 교수로 재직하며 나중에 MIT에 인공지능 연구소를 설립하게 된다. 그는 하버드에서 수학을 전공하였으며, 이 때문에 수학에 기반한 방법에 기초해 인공지능을 설계하기 위한 노력을 하였다. 한편, 그의 고등학교 1년 후배로 알려진 심리학자 프랭크 로젠블랫은 민스키와는 다르게, 뇌 과학자들의 연구를 바탕으로 인간의 뇌를 모델로 한 새로운 방법의 인공지능 기법을 제안하는데 그것이 바로 1957년 제안된 퍼셉트론(Perceptron)이다. 즉 퍼셉트론은 수학에 바탕한 인공지능의 구현이 아닌, 인간의 뇌를 본 뜬 회로를 구성해 구현하려고 시도한 새로운 방법이었다.

### 1-2 퍼셉트론의 한계

인간의 뇌를 그대로 본 뜬 전자회로라는 개념의 퍼셉트론은 세상의 관심과 주목을 받게 되고 많은 사람들의 기대를 한 몸에 받았으며, 이에 따라 로젠블랫은 상당한 연구지원 자금을 여러 곳으로부터 확보하게 된다. 로젠블랫의 이러한 약진은 그때까지 많은 주목을

받고 있었던 마빈 민스키의 상대적 소외를 야기했고 그는 당시 이러한 로젠블랫을 상당히 시기한 것으로 알려져 있다. 마빈 민스키는 1969년 한편의 책을 펴내는데, 그 내용은 바로 프랭크 로젠블랫이 만든 퍼셉트론의 근본적인 한계를 지적하는 것으로서, 퍼셉트론은 기본적으로 데이터 사이의 선형관계만을 파악할 수 있으며, 그 외의 어떠한 복잡한 문제도 해결할 수 없음을 증명하는 것이었다. 이 한 권의 책으로 인해 로젠블랫에게 쏟았던 관심은 급격히 사라지게 되었고, 동시에 인공지능에 자체에 대한 기대한 큰 실망으로 이어져 모두가 인공지능 기술에 대해 회의를 느끼게 되며 관심이 사라져 갔다.

### 1-3 1차 인공지능 빙하기

1950년대 많은 학자들은 금방이라도 사람과 같은 기계를 만들 수 있을 것처럼 큰소리를 쳤다. 1958년 허버트 사이먼은 “10년 이내 디지털 컴퓨터가 새로운 수학적론을 증명하고 발견할 것”이라고 이야기했고, 1965년에는 한발 더 나아가 “20년 이내 기계는 사람이 할 수 있는 모든 일을 해낼 것이다.”라고 예측했다. 마빈 민스키 역시 1970년에 “3~8년 이내에 평균적 인간의 지능을 가진 기계가 만들어질 것이다.”라고 큰소리쳤다. 그러나 이러한 예측은 들어 맞기는 커녕 그와 유사한 그 어떠한 업적도 보여주지 못했다. 이에 많은 투자자들은 크게 실망하고 인공지능에 투자된 모든 돈을 회수하기에 이른다. 그리하여 1974년부터 1980년까지 아무도 인공지능에 투자하려 하지 않는 인공지능의 첫 번째 빙하기가 찾아온다.

### 1-4 전문가 시스템

1차 인공지능의 빙하기는 1980년 소위 전문가 시스템(Expert System) 혹은 규칙기반 시스템(Rule-based System)의 등장과 함께 벗어나게 된다. 전문가 시스템이란 기존의 일반적인 프로그램 구현 방식과는 달리 일련의 규칙을 통해 과제를 실행하도록 한 방식을 의미한다. 즉, 일반적인 컴퓨터 프로그램은 전체적인 논리가 사전에 미리 정해지고, 이 고정된 업무 논리에 따라 프로그램이 수행되는 방식이다. 그러나 전문가 시스템은 사전에 고정된 논리가 아닌 일련의 규칙의 집합으로 이루어져 있고, 이를 수정할 때는 새로운 규칙 집합을 정의해 주기만 하면 된다.

새로 만든 보험 상품의 보험료를 계산하기 위한 프로그램을 작성해야 하는 예제를 생각해 보면 쉽다. 일반적인 프로그램 방식에서는 보험료 계산을 위해 사전에 보험료 산출을 위한 복잡한 공식을 코딩으로 입력하고 이 정해진 코드에 따라 계산이 수행된다. 보험료의 수정이 필요한 경우에는, 이 공식을 처음부터 모두 코드를 고쳐가면서 수정해야만 가능하다. 그러나 이를 전문가 시스템의 형식을 사용해 구성하면, 사전에 정해진 공식을 코드로 고정시켜 입력하는 것이 아니라, 이 공식을 이루고 있는 모든 변수와 그 변수들 사이의 관계 및 규칙을 하나의 집합으로 정의하여 입력한다. 만약 보험료의 변경이 일어날 경우에는 다시 이들 변수 사이의 관계를 새롭게 정의하게 되면 자동으로 보험료도 따라서 수정되며 보험료 변경에 간편하게 대처할 수 있게 된다.

전문가 시스템은 이러한 유연성으로 인해 많은 주목을 받았지만, 기본적으로 규칙이 어느 정도 잘 정의될 수 있는 특정 분야의 문제에만 적용 가능해서 그 응용분야가 매우 좁고, 또 효용에 비해 매우 고가에 판매한 일부 회사의 그릇된 장삿속까지 드러나 다시금 투자자들의 실망 속에 제 2의 빙하기를 초래하게 된다.

## 1-5 제 2의 빙하기와 부활

1987년부터 1993년까지 또 다시 인공지능의 빙하기가 찾아온다. 이 빙하기는 컴퓨터 성능의 비약적 발전과 더불어 기존에는 불가능하다고 여기던 몇 가지 상징적인 사건들이 발생하면서 사라진다.

그 대표적인 것이 바로 1997년 5월 11일 IBM의 딥블루(Deep Blue)가 최초로 세계 체스 챔피언을 이긴 사건이다. 당시 딥블루는 초당 2억 가지의 말의 움직임을 계산해서 사람을 이긴 것으로 알려졌다. 이와 함께 2005년에는 스탠포드 대학에서 사막에서 131마일 자율 주행을 성공했으며, 2007년에는 카네기 멜론 대학이 시카고의 도시 지역에서 55마일 거리를 자율 주행하는 데 성공하였고 이러한 흐름은 2011년 IBM의 왓슨(Watson)이 제퍼디(Jeopardy)라는 퀴즈 쇼에서 두 명의 제퍼디 챔피언을 큰 점수차로 이기면서 크나큰 주목을 이어가게 된다. 이러한 흐름은 이후 딥러닝이라는 새로운 무기와 만나 지금까지 인공지능의 역사를 새로 쓰고 있다. 딥러닝은 제6절에서 다시 자세히 알아본다.

## 2 기호주의와 연결주의

인공지능을 구현하고자 하는 인간의 욕망은 여러 가지 방식으로 시도되었지만, 그간 가장 주효했던 방식은 크게 두 가지로 분류할 수 있는데, 이는 각각 기호주의 인공지능 기법과 연결주의 인공지능 기법으로 나눌 수 있다.

### 2-1 기호주의 인공지능

기호주의의 대표적 인물은 허버트 사이먼과 마빈 민스키였다. 이들은 수학의 힘을 빌려, 인공지능을 구현할 수 있다고 생각했는데 그들이 믿은 방법은 인간의 지능이라는 것은 기호와 기호 사이에 적절한 논리 관계를 형성하여 설명할 수 있고 따라서 이를 통해 인간의 지능을 구현할 수 있다는 것이었다. 이러한 기호주의식 사고방식은 1950년대부터 1980년대 까지 인공지능의 연구방식을 주도했으며 이러한 믿음이 정점으로 구체화된 것은 앞서 잠시 설명한 1980년대의 전문가 시스템이다. 기호주의 인공지능은 “전통적(classical) 인공지능”으로도 불린다.

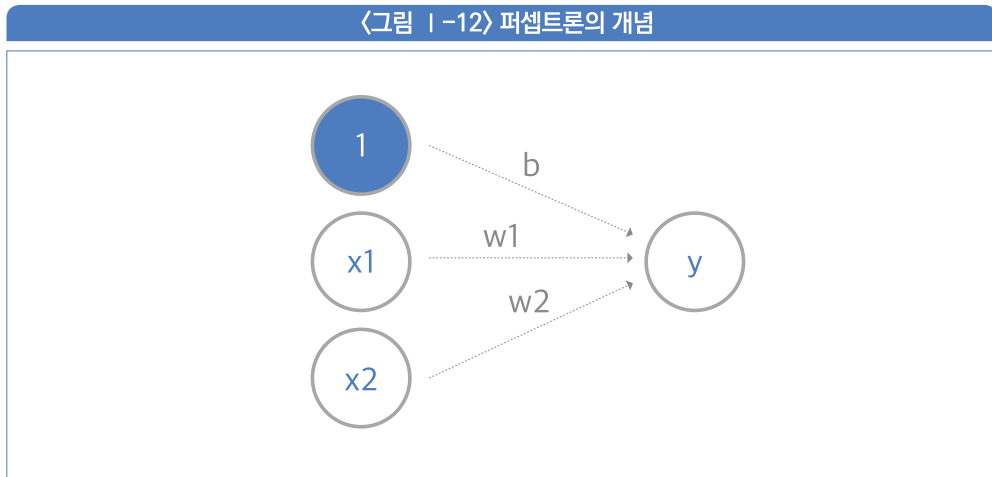
전문가 시스템은 지금도 특히 금융 분야에서는 많이 사용되고 있다. 보험사의 보험료 계산 시스템, 상품 비교 시스템 등의 영역에서는 전문가 시스템이 잘 들어맞는 분야가 많은데, 그 이유는 상품의 구성과 보험료의 계산 등은 모두 변수와 변수 사이의 관계로 설명할 수 있기 때문이다.

### 2-2 연결주의 인공지능

연결주의의 선봉은 프랭크 로젠블랫이었다. 연결주의는 뇌 과학자들의 연구에서 영감을 받은 방법으로, 인체의 뉴런과 시냅스라는 신경 조직이 전류를 흘려보내 외부의 자극을 전달하는 것을 그대로 전기 회로로 구현해 보자는 것에서 출발했다. 우리 몸은 전류가 흐르며 외부의 자극을 뉴런을 통해 전달하는데 이때 자극이 일정 이하이면 그 다음 뉴런으로 자극이 전달되지 않고, 자극이 일정 임계치 이상일 때만 그 다음 뉴런으로 지속적으로 전달한다.



1957년 프랭크 로젠블랫은 이러한 작동 기저를 그대로 흉내 내어 간단한 퍼셉트론이라는 개념을 구현했는데 그 모습은 다음 그림과 같다.



〈그림 1-12〉은 퍼셉트론의 개념도를 보여준다. 그림에는 외부의 입력 변수  $x_1 \sim x_2$ 가 회로를 통해 모여  $y$  값을 이룬다. 이  $y$  값은 다음과 같이 표현할 수 있다.

$$y = x_1 * w_1 + x_2 * w_2 + b$$

그림에서 화살표는 우리 몸의 신경 세포들에 비유할 수 있고, 원으로 표현된 것은 뉴런에 비유할 수 있다. 즉, 마치 외부 자극( $x_1, x_2$ )이 일정 가중치( $w_1, w_2$ )로 전달되었을 때, 그 전체 자극이 하나의 뉴런에 모여지는 모습을 그대로 흉내 낸 것이다. 즉, 외부 자극의 크기는 가중치 ( $w_1, w_2$ )를 통해 흉내 내며, 이 것이 충분히 클 때만  $y$  값에 의미 있는 수치로 전달된다.

프랭크 로젠블랫은 뉴런 개수에 해당하는 전기 회로 수를 더 복잡하게 구성할 수 있다면 사람과 유사한 기능을 수행할 수 있는 인공지능을 구현할 수 있을 것이라 생각했고, 이러한 연결주의적 생각은 지금까지 이어져 현재의 딥러닝으로 발전하게 된다.

인공지능이라는 말은 흔히들 영화 속의 터미네이터 같은 사람과 같은 행동을 하는 그 무엇을 연상시킨다. 1950년대 인공지능을 연구하던 많은 학자들은 스스로도 사람과 같은 기계를 만들 수 있었다고 믿었지만 그것은 한낱 몽상에 불과했다는 점을 깨닫는 데는 그리 오래 걸리지 않았다. 현대에 와서도 여전히 인공지능이라는 단어는 사람과 같은 기계를 만드는 것이라는 착각을 주지만 사실 터미네이터 같은 기계를 연구하고 있는 사람은 사실상 찾아보기 힘들다.

### 3-1 강 인공지능과 약 인공지능

강 인공지능(Strong AI)은 소설 속에 등장하는 사람과 같이 행동하고 생각할 수 있는 기계를 지칭하는 말이다. 일반적으로 인공지능은 강 인공지능처럼 잘못 알려져 있는데, 사실 강 인공지능을 연구하고 있는 사람은 거의 없으며, 그 방법도 알려진 것이 거의 없다.

우리가 흔히 인공지능이라는 단어로 쓰고 있는 여러 기술들, 예컨대 자율주행이나 의료 이미지 판독 등은 모두 강 인공지능이 아닌 약 인공지능(Weak AI)이다. 즉, 사람처럼 생각하는 기계를 만드는 것이 아니라 사람의 아주 특정영역 그중에서도 아주 지엽적인 일부분만을 겨우 흉내 내는 기술이라는 뜻이다. 터미네이터 같은 기계는 말 그대로 소설이나 영화에서만 가능하다.

학문적으로 강 인공지능은 범용 인공지능(AGI, Artificial General Intelligence)으로도 불리며 이는 현재 인공지능 학문영역의 분야에 속하지도 않는다. 현재의 학문적 인공지능 연구의 분야는 오직 약 인공지능이며, 이는 제한적 인공지능(ANI, Artificial Narrow Intelligence)이라고도 불린다.

### 3-2 인공지능의 정의

인공지능에 대한 정의는 조금씩 다를 수 있다. 여전히 사람의 지능을 구현한다거나 사람의 지능을 흉내 내는 것 등의 표현을 쓰는 사람도 적지 않지만, 아직도 사람의 지능이 어떻게 작동하는지 전혀 알지 못하는 상태에서 지능을 구현한다는 것은 적절한 표현이 아니며, 무엇보다도 인공지능에서 사용하는 기법들은 사람의 지능과는 거리가 먼 방식이다. 그런 관점에서 가장 기술 중립적이고 보편적인 인공지능의 정의는 다음과 같이 간단하고도 명료하게 내릴 수 있다.

“컴퓨터가 인간의 특정 행동을 흉내 낼 수 있게 하는 모든 기법”

여기서 명백한 점은 지능 여부와 상관없이 단순히 인간의 특정 행동을 흉내 낸다는 것으로서, 흉내를 내기 위한 ‘방법’에는 어떠한 구체적인 기술도 하지 않고 있다. 예를 들어 X-ray 사진을 보고 폐렴 여부를 판단할 때, 지금의 인공지능 기술은 의사보다 더 정확하게 판단할 수 있지만, 이것은 의사의 지능을 따라하는 것이 아니라 의사가 폐렴을 판독하는 행동을 컴퓨터 비전이라는 기술로 따라하는 것뿐이다. 이를 굳이 따라한다고 표현하는 것은 복잡한 개념을 이해하고 사물의 관찰에서부터 판독하는 의사의 판단과정과 단순히 모든 데이터를 임의의 아주 큰 수치로 변환하여 수치 상의 차이만을 찾을 수 있는 인공지능의 판단과정은 완전히 다르기 때문이다. 컴퓨터는 수치만을 이해할 뿐, 어떠한 개념도 이해시킬 수 없다.

### 3-3 인공지능의 분류

〈그림 1-13〉은 인공지능 기법의 분류를 보여주고 있다. 그림에서 보는 것처럼 인공지능 기법은 그보다 더 작은 범위인 머신러닝 기법과 딥러닝으로 보다 세분화할 수 있다.

머신러닝은 인공지능 기법의 한 종류이지만 최근에는

〈그림 1-13〉 인공지능의 분류



거의 대부분의 인공지능 기법이 머신러닝을 이용하기 때문에 인공지능이면 곧 머신러닝이라 할 정도로 보편화되고 있다. 머신러닝의 가장 큰 특징은 데이터에서 유의미한 패턴을 추출할 때 기계 스스로 학습하는 알고리즘을 사용한다는 것이다. 즉, 인공지능의 여러 알고리즘 중 데이터에 숨겨진 패턴을 찾아내기 위해 스스로 학습할 수 있는 모든 알고리즘은 머신러닝이라 부르는 것이다. 머신러닝에 속하지 않는 대표적인 인공지능 기법은 앞서 잠시 언급한 전문가 시스템이다.

한편 최근에 각광받고 있는 딥러닝 역시 머신러닝의 영역이다. 즉, 딥러닝 역시 데이터 내의 숨겨진 패턴을 찾기 위해 스스로 학습할 수 있는 알고리즘이다. 다만 이때 심층 신경망이라 불리는 특정 아키텍처를 사용한 머신러닝 기법을 특별히 딥러닝이라고 부른다. 심층 신경망이란 인간의 뇌를 흉내 낸 인공 신경망이라는 구조를 매우 여러 겹 쌓은 것을 의미하며, 이때 여러 겹을 의미하는 심층의 영문 표현이 바로 딥(deep)이다. 머신러닝과 딥러닝에 대한 보다 자세한 설명은 제6절에서 하기로 한다.

인공지능을 통해 특정 문제를 해결하는 방법은 크게 3가지로 분류할 수 있다. 이는 대체로 주어진 문제의 다양한 성질에 따른 분류방법으로서 각 문제의 고유한 특성으로 인해 어느 한 가지 방법으로는 이 모두를 해결할 수 없기 때문이다. 이 절에서는 각 방법에 대해 자세히 살펴보도록 하자.

## 1 지도학습

지도학습은 이미 정답을 알고 있는 데이터를 사용해, 숨겨진 규칙을 찾아내고 이를 일반화하려는 방법이다. 예를 들어 은행의 대출 데이터를 생각해 보자. 은행은 그간 여러 사람들에게 대출을 실행해서, 그 결과 즉, 최종적으로 특정 대출자가 결과적으로 변제를 했는지 여부에 대한 결과 데이터를 가지고 있다. 이때, 대출한 사람의 특성을 나타내는 '특징'을 X라는 변수로 나타내고, 최종적으로 이 특징인이 변제를 완료했는지를 나타내는 변수를 Y로 표현하면, 우리는 다음의 관계식을 얻을 수 있다.

$$X \rightarrow Y$$

이때 특정인의 특성을 나타내는 독립변수<sup>13)</sup> X에 대해, 이 사람이 결과적으로 대출금을 변제했는지 그 결과를 나타내는 값 Y는 이미 은행이 획득하여 가지고 있는 상태이다. X는 일반적으로, 특정인의 나이, 성별, 직업, 연봉수준, 결혼상태 등의 다양한 값을 조합해 사용할

13) 이를 특징(feature)이라고도 부른다.

수 있을 것이며 이를 수치화하여 나타낸 데이터이다. 이 데이터는 일반적으로 다음과 같이 나타낼 수 있다.

(나이, 성별, 직업, 연봉, 결혼상태, 대출금) → Y

즉, 가상의 특징인, 예컨대 30대 남성인 홍길동 씨의 직업은 사무직이고, 중간(中) 정도의 연봉을 받으며, 현재 결혼 상태인데, 1억의 대출금을 받았다고 하자. 이 사람이 결과적으로 자신의 대출금을 충실히 잘 변제했다고 가정해 보자. 대출금을 변제했을 경우를 1이라고 표기하고, 이를 변제하지 않았을 경우를 0이라고 나타낸다면, 우리는 홍길동 씨의 특정 경우를 다음과 같이 나타낼 수 있을 것이다.<sup>14)</sup>

X = (30, 남, 사무직, 中, 기혼, 1억) → Y = 1

은행은 기존에 대출을 집행했던 수많은 사람들에게 대한 이러한 데이터를 이미 가지고 있을 것이므로 이제 머신러닝 알고리즘을 사용해 어떤 특징의 사람일 경우 과연 돈을 충실히 변제했는지 그 규칙을 찾아내는 것을 시도해 볼 수 있다. 일단 규칙을 찾아냈다면, 이제 새로운 고객 김철수 씨가 은행에 신규 대출을 신청할 경우 이 사람이 궁극적으로 대출금을 갚을지 예측할 수 있게 된다. 즉 김철수 씨가 40대이며, 남성이고 운동선수이며, 연봉이 높고, 미혼이며 2억을 대출하려 한다면 찾아낸 규칙에 이 특징을 대입하여 그 결과를 예측해 볼 수 있는 것이다.<sup>15)</sup>

X = (40, 남, 운동선수, 高, 미혼, 2억) → Y = ?

이처럼 이미 정답을 알고 있는 데이터로부터 규칙을 찾아내어 새로운 데이터를 예측하려는 기법을 지도학습이라고 한다.

14) 실제로는 모든 특징들은 글자가 아니라 수치화하는 변환작업을 거쳐야 한다.

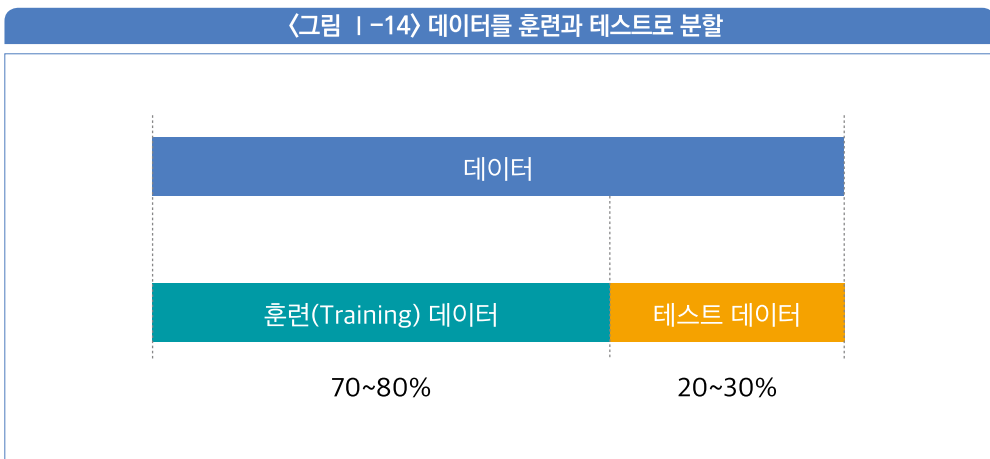
15) 실제로는 대출변제 유무인 1,0 대신 변제를 할 확률값을 구하도록 식을 구성하는 것이 훨씬 더 유용할 것이다.

## 1-1 훈련 데이터와 테스트 데이터

인공지능의 목적은 미래의 예측을 위한 것이다. 따라서 현재 데이터를 분석하는 이유는 분석결과로 얻은 규칙이 ‘일반화’되어 앞으로 보게 될 다른 새로운 데이터<sup>16)</sup>에도 잘 적용되기를 바라는 것이다.

그러나, 규칙을 찾기 위해 사용하는 과거의 데이터가 전체 모집단을 얼마나 충실히 대표하고 있을지는 아무도 모른다. 이 때문에 현재 손에 주어진 과거의 데이터가 전체 모집단을 완벽하고도 충실히 대표할 것이라 믿는 것은 좋지 않은 방법이며, 이 때문에 데이터에 대해 적당한 ‘의심’을 하는 것이 좋다.

데이터를 적당히 의심하며 모델을 만들기 위해서 가장 흔히 쓰는 방법은 현재 주어진 데이터를 두 개의 그룹으로 나눈 다음 그중 하나만을 이용해 모델을 훈련시키고 나머지 하나는 이 모델이 잘 일반화될 수 있는지 검정해 보는 테스트 용으로 사용하는 것이다. 다음 그림을 보자.



16) 이를 통상 낯선(unseen) 데이터라고 한다.

〈그림 1-14〉는 주어진 과거 경험 데이터를 훈련 데이터와 테스트 데이터로 나누고 있는 모습을 보여준다. 대개 훈련 데이터는 전체 데이터의 70%~80%, 테스트 데이터는 전체 데이터의 20~30%를 활용한다.

훈련 데이터는 말 그대로 훈련에만 사용되며 이 과정에 절대 테스트 데이터가 이용되어서는 안 된다. 모델은 오로지 훈련 데이터에 있는 정보만을 사용하여 규칙을 찾아내야 한다. 이렇게 찾아낸 규칙은 테스트 데이터를 대상으로 그 성능을 평가하게 되는데, 만약 훈련 데이터에서의 성능은 매우 좋았지만 테스트 데이터의 성능이 좋지 않다면 이는 모델이 찾아낸 규칙이 일반화되기 힘들다는 의미가 된다.

물론, 데이터가 수십만 개 정도로 충분히 많다면 굳이 이렇게 데이터를 쪼개서 사용하지 않아도 된다. 표본 데이터의 개수에 비례하여 점점 더 전체 모집단을 충실히 잘 대표할 수 있음은 자명할 것이기 때문이다. 이는 바꿔 말하면 표본 데이터가 적을수록 규칙을 일반화하기 힘들다는 의미도 된다.

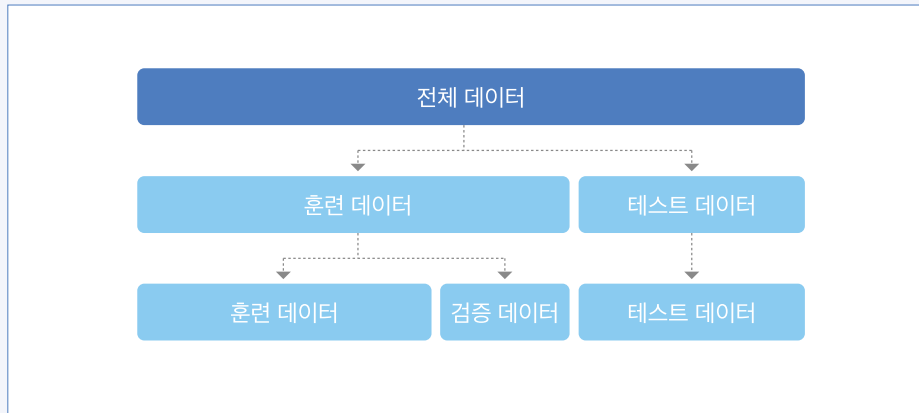
#### MEMO

머신러닝의 핵심은 모델을 정의하는 매개변수를 알고리즘이 스스로 학습하며 찾아낸다는 것이다. 그래서 모델 정의에 최대한 사람의 개입을 줄이지만, 완전히 사람을 배제할 수는 없다. 따라서 사람이 직접 모델의 구조에 개입되는 상황이 벌어지는데, 이때 사람이 직접 지정해주는 매개변수를 특별히 초매개변수(Hyper Parameter)라고 부른다. 초매개변수의 적절성에 따라 모델 성능은 크게 달라질 수 있는데 이때 역시 적절한 초매개변수를 찾기 위해 별도의 데이터를 사용하기도 한다. 이 경우 데이터는 두 개가 아닌 세 조각으로 나뉘어 지는데, 각각 훈련 데이터, 테스트 데이터, 검증 데이터이다.

가장 일반적인 방법은 우선 데이터를 훈련과 테스트 두 조각으로 나눈 다음, 훈련 데이터를 다시 훈련과 검증 데이터로 나누는 방식이다. 모델의 초매개변수의 적절성은 검증 데이터를 통해 확인한다. 이후 최종 성능은 앞서와 마찬가지로 테스트 데이터를 이용해서 수행한다.



〈그림 1-15〉 전체 데이터, 훈련 데이터, 테스트 데이터, 검증 데이터



## 1-2 과적합과 과소적합

데이터로 모델을 훈련시킬 때에는 통상 두 가지 큰 오류를 범할 수 있는데, 하나는 과적합(Over-Fitting)이고 다른 하나는 과소적합(Under-Fitting)이다. 과적합이란, 주어진 훈련 데이터에만 너무 충실히 의존하여 모델이 훈련 데이터를 대상으로 한 성능 평가에는 매우 우수한 결과를 나타내지만, 테스트 데이터를 상대로 한 성능 평가에는 좋지 않는 결과를 나타나는 경우로서, 모델이 찾아낸 규칙이 일반화되기 힘들다는 의미가 된다.

또 다른 경우는 훈련 데이터에 숨겨진 규칙을 충분히 찾아내지 못한 경우로서, 이때는 훈련 데이터의 성능과 테스트 데이터의 성능이 모두 좋지 않게 나타난다.

최적의 모델을 생성하기 위해서는 과적합과 과소적합을 모두 피해야 하며, 따라서 훈련 데이터와 테스트 데이터 모두에서 좋은 성능을 보여주어야만 비로소 규칙이 일반화될 수 있다는 가정을 해 볼 수 있다. 통상 모델을 기술하는 매개변수의 수가 증가할수록 과적합에 빠지기 쉽다.

## 2 비지도 학습

특정 과제에서는 원래 정답이라는 것 자체가 존재하지 않는 경우가 적지 않다. 혹은 정답이 존재할 수는 있지만 사전적으로는 그 정답을 알 수 없거나, 혹은 정답을 기록해 둔 데이터를 가지고 있지 않은 경우에는 통상 지도학습에서 사용하는  $X \rightarrow Y$ 라는 알고리즘 기법을 그대로 적용하는 것이 용이하지 않으므로, 다른 알고리즘 기법이 필요하다. 이처럼, 정답이 없는 문제 혹은 정답을 알 수 없는 문제를 해결하는 방법에 특화되어 사용되는 것을 비지도 학습이라고 한다.

정답을 알 수 없는 대표적인 문제 중 하나는 회사에서 마케팅을 위해 고객을 세분화하는 경우이다. 이때 고객을 비슷한 유형의 그룹으로 나눌 때, 과연 몇 개의 그룹으로 나누는 게 가장 좋을 것인지, 어떤 기준으로 나누는 것이 최적인지에 대해 정해진 정답이 있는 것은 아니다.

비지도 학습의 전형적인 방법 중 하나는 데이터를 유사한 특성을 가진 것끼리 서로 그룹화하는 군집화<sup>17)</sup> 방법이다. 이때 몇 개로 군집화할 것인지, 또는 어떤 기준으로 군집화 할 것인지는 대개 사람이 개입하여 지정해 주거나 몇 가지 다른 기준을 만들어 시도해 본 다음 어느 것이 더 적절한지 판단하여 그중 최적의 기준을 선택하기도 한다. 이때 각각의 군집을 클러스터(cluster)라고 부른다.

한편, 비지도 학습은 원래는 정답이 있는 문제이지만, 정답이 기록된 데이터를 가지고 있지 않은 경우에도 활용되기도 한다. 예를 들어, 주어진 이미지를 보고 고양이인지 개인지 판단하는 과제를 생각해 보자. 이러한 모델의 학습에는 지도학습 기법이 가장 이상적이지만 지도학습을 하기 위해서는 모든 이미지에 고양이 그림인지 개의 그림인지 사람이 일일이 개입하여 판단하고 그 정답에 해당하는 레이블(label)을 입력해 두어야만 모델 훈련이 가능하다. 그러나 이러한 데이터 준비 작업에는 많은 인건비가 소요될 수 있다.

---

17) 이를 영어로는 클러스터링(Clustering)이라고 한다.

예컨대 이미지가 100만 장이라면, 이 작업에 엄청난 비용과 시간이 소요될 수 있을 것이다. 이 경우 비용 절감을 위해서 비지도 학습을 활용해 볼 수 있다. 즉, 비지도 학습의 군집화를 통해 전체 이미지를 두 부류로 나누도록 한 다음 이를 기반으로 지도학습을 수행할 수 있는 것이다. 이때, 군집화에 사용되는 알고리즘은 개와 고양이를 정확히 구분하는 것이 아니라 그저 수치로 변환된 이미지를 서로 유사한 수치끼리 묶어서 분류하는 것에 불과하다. 따라서 사람이 직접 분류하여 정답을 부여하는 것에 비하면 데이터의 정확도가 현저히 떨어질 것은 자명하다.

한편 이렇게 두 가지 그룹의 군집화한 결과를 눈대중으로 본다음 개가 더 많아 보이는 그룹에는 일괄적으로 개라고 한꺼번에 레이블을 자동으로 붙이고, 다른 그룹에는 고양이라고 레이블을 자동으로 달면 이제 정답이 있는 데이터로 변환된 셈이므로 지도학습이 가능하다. 이런식으로 데이터를 준비하는 데는 사람이 개입되어 수작업을 하는 경우에 비해 전체 작업이 매우 간단해 진다. 물론, 상당히 부정확할 수도 있는 이 데이터를 사용해 지도학습을 통해 훈련시킨 모델은 정확한 데이터에 기반하여 훈련시킨 모델에 비해 그 성능이 상당히 떨어질 것은 당연하지만 전체 비용 측면에서는 어느 정도 합리적일 수 있다.

### 3 강화학습

앞서 살펴본 지도학습이나 비지도 학습 기법을 모두 동원하더라도 해결할 수 없는 문제 영역도 존재하는데 바로 강화학습의 영역이다. 대표적인 예로는 바둑과 자율 주행, 로봇의 움직임 제어 등이 있다.

바둑의 한수 한수는 어떤 정답이 정해져 있는 것이 아니다. 예컨대 특정 지점에 검은 돌을 두었을 때 그 다음 흰 돌의 위치는 고정된 절대 정답을 가지고 있지 않다. 모든 프로 기사가 항상 같은 위치에 돌을 두지는 않을 것이기 때문이다.

한편, 특정 시점에서 어느 위치에 일단 돌을 두고 나면, 그 결정으로 인해 그 다음 모든 수에 영향을 미치게 된다. 이는 자율주행도 유사한데, 어느 특정 시점에서 핸들을 왼쪽으로 꺾을지 오른쪽으로 꺾을지는 정답이 정해진 것이 아니며, 그 순간 어떻게 행동했는지에 따라 그 다음 행동에 영향을 미친다는 특징을 가지고 있다. 그리고 앞으로 주행하면서 맞이하게 될 도로의 구체적인 상황이 어떻게 펼쳐질지는 아무도 모른다.

이처럼 정답도 없고, 현재 환경을 완전히 이해할 수 없는 상태 즉, 미지의 상태에서 지속적으로 연속적으로 ‘바람직한’ 결정을 내려야 하는 문제 영역은 앞서의 지도학습이나 비지도 학습법으로는 해결할 수 없다. 그래서 나온 완전히 새로운 알고리즘 구조가 바로 강화학습이다.

〈그림 1-16〉 일반적인 강화학습 모델



〈그림 1-16〉은 일반적인 강화학습 모델을 보여준다. 이 모델에서는 두 가지 중요한 개체가 존재하는데 바로 에이전트와 환경이다. 이 두 개체 사이의 상호작용은 보상함수를 통해 기술한다.

먼저 에이전트는 주어진 환경에서 특정 행동을 할 수 있는 개체로서 사전에 정해진 특정한 행동을 취할 수 있다. 예를 들어 자율 주행을 강화학습으로 모델링하는 경우라면, 에이전트의

행동은 {핸들을 왼쪽으로 이동, 핸들을 오른쪽으로 이동, 브레이크 밟기, 액셀 밟기} 등의 집합으로 정의할 수 있을 것이다. 바둑의 경우라면 특정 좌표에 바둑 돌을 두는 것으로 에이전트의 행동을 정할 수 있으며, 미로 찾기에는 {동으로 이동, 서로 이동, 남으로 이동, 북으로 이동}으로 정의할 수 있을 것이다.

환경은 이 에이전트의 특정 행동에 의해 영향을 받게 된다. 이 에이전트의 행동은 때로는 환경에 긍정적 영향을 미치겠지만 경우에 따라서는 부정적 영향을 미칠 수 있을 것이다. 예를 들어, 자율 주행에서 핸들을 잘못 꺾어 충돌이 야기된다면 이는 환경에 좋지 않은 영향을 끼친 것이므로 지양해야 할 것이다.

이때 에이전트의 특정 행동으로 인해 환경에 부정적 영향을 미쳤는지 긍정적 영향을 미쳤는지 판단하고, 이를 점수화하여 피드백을 주는 것이 바로 보상함수이다. 즉 보상함수의 값이 클수록 환경에 긍정적인 영향을 끼치는 것이므로 에이전트는 그 방향으로 훈련되어야 하고, 이 점수가 음수라면 환경에 부정적인 영향을 끼친 것이므로 그러한 행동은 지양되어야 한다. 에이전트는 끊임없이 어떤 행동을 하고, 특정 행동을 할 때마다 보상함수는 환경에 생긴 변화를 점수화하여 에이전트에 통보한다. 에이전트는 이 점수가 극대화되는 방향으로 끊임없이 자기의 행동을 보정해 나가면 비로소 환경에 바람직한 방향으로 영향을 끼치는 쪽으로만 훈련되어 가는 것이다.

강화학습은 최근에 더욱 조명을 받고 있는데, 그 이유는 학습을 위해 준비해야 할 데이터가 필요 없기 때문이다. 앞서 잠시 살펴본 것처럼 지도학습의 경우 훈련을 하기 위한 '정답이 기록된' 데이터를 사전에 확보해야만 하는데 이러한 작업에는 상당한 비용이 소요된다. 많은 기업들이 지도학습을 도입할 때 겪는 가장 큰 장애물 중 하나가 바로 데이터의 수집비용이다.

#### MEMO

강화학습은 지도학습과 혼용하여 사용되기도 한다. 예를 들어 자율비행 문제 자체는 강화학습에 적합하다. 그러나 최정에 파일럿의 비행 모습을 모델로 삼아 지도학습과 병행하는 것이 가능하다. 이런 방법을 쓰면 훨씬 더 빨리 모델을 훈련시킬 수 있다. 앞서 개와 고양이의 구분에 지도학습과 비지도 학습을 혼합해서 사용하는 예처럼 하나의 학습법이 반드시 고립된 형태로 사용될 필요는 없다. 더 효율을 높일 수 있고 비용측면에서 더 유리하다면 다양한 조합을 통해 모델을 훈련시킬 수 있다.

또한 정답을 정의하기에 따라 강화학습도 지도학습의 영역이 되기도 한다. 예를 들어 바둑의 정답을 모든 프로기사의 대국 방법이 아닌 특정 프로 기사의 대국 방법으로 국한하면, 그 프로기사의 기보를 사용해 지도학습을 할 수 있는 것이다. 실제로 이세돌 씨를 이긴 알파고는 프로 기사들의 기보를 바탕으로 지도학습을 통해 훈련시킨 것이다. 그 후 알파고-제로 모델은 기보를 바탕으로 한 지도학습이 아닌 강화학습을 통해 훈련시킨 것이다.

## 4 판별적 모델과 생성적 모델

앞서 살펴본 인공지능 모델은 주로 주어진 데이터에 존재하는 어떤 규칙을 찾아내어 일반화한 다음 미래에 새로운 데이터를 보고 예측을 수행하기 위한 것이었다. 이러한 모델을 판별적 모델이라고 한다. 최근 인공지능 분야에는 이러한 전통적인 모델과 정반대의 응용도 활발히 연구되고 있는데 바로 생성적 모델이 그것이다.

〈그림 1-17〉은 구글이 생성적 모델을 사용해 만들어 낸 가상의 여성 이미지이다. 이 이미지는 실존하는 인물이 아닌 컴퓨터가 합성으로 만들어 낸 가상의 인물인 것이다. 이처럼 원래 존재하지 않는 이미지 등을 컴퓨터 인공지능 기술을 이용해 만들어 내는 기법을 생성적 모델이라고 한다.

〈그림 1-17〉  
구글이 생성적 모델로  
만들어낸 가상의 여성 인물

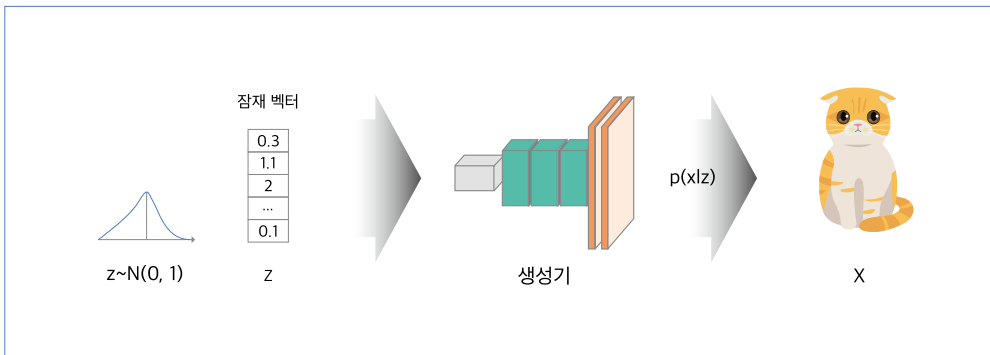


출처: 구글 합성이미지  
(<https://generated.photos/>)

## 4-1 생성적 모델

생성적 모델은 판별적 모델과 달리, 무에서 유를 창조하는 것에 비유할 수 있다. 생성적 모델은 랜덤 분포에서 시작해 어떤 의미있는 목표물을 생성한다. 예를 들어, 아무것도 없는 잡음에서 출발해 고양이 이미지를 생성할 수도 있다.

〈그림 1-18〉 생성모델

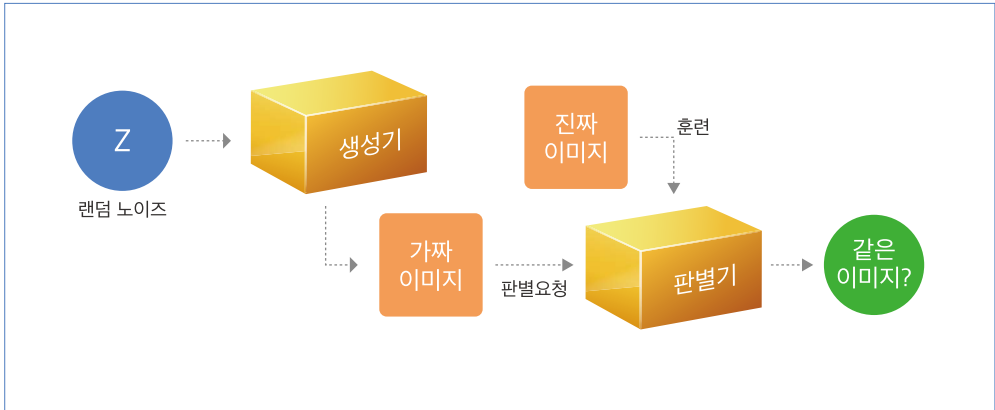


〈그림 1-18〉은 생성모델을 사용해 고양이 그림을 창조해 내는 예시를 보여주고 있다. 그림에서 보는 것처럼 임의의 분포에서 출발해 최종적으로 고양이 이미지를 생성해 낸다. 이러한 생성모델에는 몇 가지 다른 알고리즘 기법이 사용되는데, 그중 최근 가장 주목을 받고 있는 적대적 생성 신경망이라 불리는 GAN(Generative Adversarial Network) 기법만 간단히 살펴해보도록 하자.

## 4-2 GAN

GAN은 적대적 생성 신경망이라 불리는 대표적인 생성적 모델이다. 이 모델은 다음 〈그림 1-19〉와 같이 구성된다.

〈그림 1-19〉 전형적인 GAN 모델



〈그림 1-19〉는 이미지 생성에 사용되는 전형적인 GAN 모델을 보여준다. 이 모델의 목표는 원하는 이미지를 랜덤 노이즈로부터 생성해 내는 것이다.

GAN 모델은 생성기와 판별기라는 두 가지 모듈로 구성되어 있다. 생성기는 랜덤 노이즈로부터 임의의 가짜 이미지를 생성해 내는 역할을 하며, 이렇게 생성한 이미지는 판별기로 보낸다. 판별기는 진짜 이미지를 사용해 훈련을 받으며, 생성기로부터 전달받은 이미지가 가짜인지 진짜인지 ‘판별해 내는’ 임무를 맡는다.

전체 과정은 생성기가 만들어낸 이미지를 전달받은 판별기가 더 이상 가짜인지 진짜인지 구분하지 못할 때까지 생성기가 반복적으로 가짜 이미지를 변경해간다. 어느 순간 결국 판별기가 이 가짜 이미지를 더 이상 가짜인지 구분하지 못하는 정도에 이르면, 드디어 원하는 목적을 달성한 것이 된다.



### 4-3 생성적 모델의 응용

〈그림 1-20〉 생성적 모델을 사용한 이미지 복원



#### 가. 이미지 재생(In-painting)

〈그림 1-20〉은 생성적 모델을 사용하는 응용프로그램 중 대표적인 이미지 재생을 보여주고 있다.

그림의 왼쪽은 원본이미지가 훼손되어 있는 것을 보여준다. 이미지 재생 기술은 이처럼 원본이 훼손되어 있는 이미지를 자연스럽게 복원하는 기술로서 우측의 그림에서 보는 것처럼 원본 영상을 최대한 충실하게 재현해 낸다.

생성적 모델은 이처럼 훼손된 이미지 부분을 복원하거나, 고 문서의 훼손된 부분의 재생, 또 흑백 영화를 컬러 영화로 변환하는 등에 많이 이용될 수 있다.

#### 나. 초고해상도(Super Resolution)

앞서 살펴본 이미지 재생과 유사한 원리로, 해상도가 떨어지는 원본 이미지를 더욱 해상도가 높은 고해상도의 이미지로 변환하는데 사용되는 GAN 기술을 초고해상도(Super Resolution)라고 한다.

초고해상도 기술을 이용하면 여러 번 확대함으로써 그 가독성이 떨어지는 이미지나 혹은

위성사진으로 찍힌 다소 불명확한 이미지 속의 글자나 인물 등을 또렷이 재생할 수 있다. 또한 CCTV 영상 속의 흐릿한 인물 사진도 보강하여 더욱 또렷한 형태로 만들 수 있으며 이러한 기술은 치안 등에도 많이 응용될 수 있다.

#### 다. 딥페이크(Deep Fake)

한편, 영상 속의 어떤 인물을 다른 사람으로 합성하는 데에도 GAN 기술이 사용되는데 이를 딥페이크(Deep Fake)라고 부른다.

딥페이크는 원본 이미지나 원본 동영상의 특정 인물을 다른 얼굴로 대체하는 기술에 관한 것으로서 주로 영화 등에서 많이 응용되기도 한다.

그러나 최근에는 이 기술이 악용되어 금융사기, 가짜뉴스, 위조 등 여러 범죄에 악용되기도 한다. 예를 들어, 유명 인물의 얼굴에 딥페이크 기술을 적용하여 만든 여러 패러디나 우스개 합성 영상물을 이전 어렵지 않게 인터넷에서 찾아볼 수 있을 정도로 딥페이크 기술은 일반인에게도 많이 퍼져 있을 정도이다. 인터넷에는 트럼프나 오바마 혹은 유명 연예인 등이 유창한 한국어로 우스개 소리를 하는 장면이 자연스럽게 합성된 딥페이크 동영상도 어렵지 않게 찾아 볼 수 있다.

#### 라. 기타 창작 영역

GAN 기술의 또 다른 응용으로는 창작영역을 들 수 있다. 최근 GAN 기술은 작곡, 그림 등 미술품을 생성하는 데 이용된다. 램브란트 등 사망한 거장의 그림 풍으로 새로운 그림을 그려내는 인공지능, 새로운 창작 소설을 써 내려가는 인공지능 등도 쉽게 볼 수 있다.

인공지능으로 생성된 그림을 과연 창작물로 볼 수 있는가에 대한 찬반 논의는 뜨겁지만, 뉴욕의 크리스티 경매장에서는 프랑스의 예술 집단 오비어스(Obvious)가 인공지능으로 그린 그림인 “벨라미 가문의 에드몽 벨라미”라는 작품이 43만 2500달러(한화 약 4억 9천만원)에 낙찰되기도 했다. 공교롭게도 같은 날 거래된 파블로 피카소의 그림 ‘젊은 여성의 초상’과 동일한 가격에 낙찰되어 더욱 화제가 되었다. 컴퓨터로 생성한 이미지가 피카소의 진품 작품과 동일한 대접을 받은 셈이다.

## 5 인공지능 비즈니스 사례

지금까지 인공지능을 사용해 과제를 해결하는 몇 가지 방법에 대해 간략하게 살펴보았다. 이제 이러한 인공지능 기술이 실제 비즈니스에는 어떻게 활용되고 있는지 그 대표적인 사례와 유형 몇 가지만 간단히 살펴보도록 하자.

### 5-1 이미지 인식 기반 서비스

이미지 인식분야는 인공지능 중에서도 가장 앞서있는 것 중 하나이다. 이미지 인식 분야는 2015년을 기점으로 이미 사람의 이미지 분류 수준을 뛰어넘었고, 이제 다양한 분야에서 활용되고 있다.

금융 비즈니스에서 이미지 인식이 활용되고 있거나 연구되고 있는 사례는 매우 많고 다양하다.

#### 가. 신분증 인식

현재 금융권에서 가장 보편적으로 사용되고 있는 이미지 인식기술 중 하나는 원격 인증을 통한 비대면 신분증 인식이다. 최근 비대면 계좌 개설이 허용되고 따라서 모바일이나 인터넷을 통해 금융 계좌를 개설할 때 본인 인증을 위해 신분증을 카메라 등을 사용해 이미지화 한후 그 내용을 인공지능 기술로 인식하여, 본인인지의 여부를 원격으로 알아보는 서비스가 매우 보편화되어 있다.

#### 나. 위조지폐 감별

이미지 인식기술은 위조지폐의 감별에도 활용되고 있다. 우리나라의 경우 국립과학수사 연구원이 한국은행과 함께 딥러닝을 활용하여 위조지폐를 감별하는 기술을 연구하였으며, 현재 이를 시중에 보급하고 있다.

#### 다. 안면인식

점포의 CCTV에 안면인식이 가능한 이미지 인식기술을 적용하려는 사례도 많이 연구되고 있다. 이러한 기능이 부착되면, 점포에 들어서는 고객이 기존고객인지 신규고객인지 판단하고, 기존고객일 경우에는 과거 거래 이력을 창구 직원의 단말기에게 자동으로 보여주어, 서비스 시간 단축과 품질 향상을 꾀할 수 있다. 그러나, 이러한 안면인식 기능은 고객의 개인정보 보호 문제와 함께 여러 가지 도덕적 윤리적 문제와 연관될 수 있으므로 신중히 도입해야 한다.

#### 라. 각종 문서의 인식 및 해석

각종 문서를 이미지로 읽은 다음, 그 내용을 자동으로 추출하여 문맥을 파악하는 기술도 보편화되고 있다. 대출의 신청, 보험의 청약, 보험금의 청구 등 다양한 서류에서 이미지 인식기술이 적용되고 있고 이는 자연어를 처리할 수 있는 기능과 결합되어 문서의 문맥을 파악하는 데 사용된다.

또한, 조금 더 진보된 기술로서 문서 인식기술과 함께 자연어 처리 기술을 병합하면 문서 속에 적힌 내용까지도 자동으로 추출하여 응용할 수 있게 된다. 예컨대 각종 기업들이 주기적으로 생성하는 문서인 재무제표를 비롯하여 시중의 경제 기사들 등의 문맥을 자동으로 추출하려는 시도도 많이 이어지고 있다.

### 5-2 고객 신용도 분석

고객의 프로파일에 기반해 대출 변제 가능성을 확률로 구하는 시도는 꾸준히 이어져 오고 있다. 고객 프로파일에서 규칙을 찾아내어 이를 대출 변제 가능성으로 변환하는 작업은 기본적인 머신러닝 기법에 잘 들어맞고, 또한 이 부분의 정확도를 개선한다면 은행의 수익구조가 크게 좋아진다는 점을 감안할 때 이 분야에 많은 시도가 이어지고 있는 것은 어쩌면 당연하다. 우리나라는 현재 단 3개의 업체가 개인 신용조회 시장을 독점하고 있고, 금융기관은 전적으로 이 3개 기관의 신용조회 점수에 기반해 대출을 시행하고 있다. 기업의 신용도 역시 다섯 개 정도의 업체가 독점적으로 평가하고 있다.

그러나, 이러한 신용등급 평가는 대부분 과거의 금융거래 이력에만 종속되거나 혹은 기업의 재무적 요소만 이용하여 평가되고 있으므로 금융거래 내역이 상대적으로 적은 고객이나 신규 고객은 정확한 신용등급평가를 할 수 없고, 또한 신용등급을 좌우 할 수 있는 다른 복합적인 요소들은 전혀 고려하지 못하는 단점이 있다.

또한 기업의 입장에서도, 새로 시작된 스타트업 기업체나 잠재 기술 가치가 큰 업종의 기업, 특허권을 보유하고 있는 기업 등 재무적이 아닌 요소에서 큰 가치를 가지고 있는 기업들에 대해서는 제대로 평가가 이루어지지 못하고 있다. 이 때문에 대부분의 개인 사업자나 소상공인들은 신용평가가 이루어지지 않고 거의 담보대출에만 의존하여 기업자금을 대출받고 있는 실정이다.

딥러닝을 이용해 다양한 고객 프로파일 정보로부터 보다 정확한 신용등급평가를 얻으려는 시도가 다방면에서 이루어지고 있으므로 조만간 다양한 척도가 시장에 등장할 것을 기대해 본다.

### 5-3 이상 거래 탐지 서비스

우리나라의 금융기관에는 정해진 법령에 따라 일정 조건에 부합하는 금융거래가 발생할 경우 이를 모두 금융정보분석원(FIU)에 보고할 의무가 부과되어 있다. 예를 들어 CTR(Cash Transaction Report)이라는 의무는 고액거래를 보고할 의무를 말하는데, 우리나라 특정 금융거래정보의 보고 및 이용 등에 관한 법률(이하 특정금융정보법이라 칭한다)에는 일정 이상의 고액 거래는 30일 이내에 FIU에 보고하도록 명하고 있는데 이 금액은 애초 2006년 1월 18일에 5천만원으로 규정되어 있던 것이 점차적으로 낮아져 2019년 4월부터는 1천만원이 넘는 거래부터 보고하도록 강제하고 있다. 이처럼 규칙이 잘 정해져 있는 CTR과 달리 금융기관에는 자율적으로 이상 거래를 탐지하고 보고할 의무인 STR(Suspicious Transaction Report)도 주어진다.

STR은 법령 등에 의해 정해진 규칙이 없으므로, 각 금융기관이 각자의 판단에 의해 자율적으로 걸러내는 수밖에 없다. 이 때문에 여러 다양한 인공지능 알고리즘이 STR의 탐지에 사용되고 있다. 대표적인 유형은 1) 전혀 거래가 없던 곳으로의 송금 2) 같은 금액의 반복 송금 3) 전혀 거래가 없던 지역에서의 금융거래 4) 빈번하지 않은 새벽 시간대의 거래 등이 있다. 이 밖에도 거래 패턴을 분석하여, 과거 이력과 비교할 때 통상적이지 않은 형태의 거래가 일어난 경우라면 이를 의심해 볼 수 있다.

#### 5-4 로보 어드바이저

로보 어드바이저(Robo-Advisor)는 로봇(Robot)과 조언을 하다는 영어인 어드바이저(Advisor)가 합쳐진 합성어이다. 이 프로그램의 주 목적은 전문 투자자문가들이 대체로 적지 않은 수수료를 받고 수행하던 고객의 투자 포트폴리오자문을 컴퓨터 알고리즘을 통해 수행함으로써 서비스 수수료를 저렴하게 낮출 수 있다는 취지로서, 다양한 분야에서 고객 맞춤형 서비스를 제공하는 것을 그 목적으로 하고 있다.

최근에는 다양한 인공지능 기법을 도입하여 전문 투자자문가보다 더 나은 성능을 제공한다고 주장하는 업체도 등장했지만, 아직까지 명확한 성능을 나타낸다고 볼만한 확실한 근거는 없다.

한편 단순한 포트폴리오의 재구성을 넘어 특정 자산의 매수 시점과 매도 시점, 그리고 매수 가격, 매도 가격 등도 자동으로 추출하는 기능 역시 활발히 연구되고 있다.

로보 어드바이저는 사람의 개입 정도에 따라 몇 가지 유형으로 나눌 수 있는데 사람이 전혀 개입되지 않고 완전히 컴퓨터의 알고리즘에만 의존하는 극단적인 유형부터 대부분 사람이 관여하여 판단하되, 로보어드바이저는 그저 단순 보조지표로서만 사용하는 단계에 이르기까지 다양하게 세분화할 수 있다.

로보 어드바이저의 일반적인 작동 방식은 먼저 고객이 자신의 맞춤형 서비스를 위해 정보를

입력하면, 이 정보를 분석하여 다양한 이론과 알고리즘을 적용하여 자산의 적절한 배분과 균형에 대한 서비스를 제공하는 것이다. 자산의 배분 역시 나이에 따라 더 안정적인 자산으로 균형을 자동으로 맞추는 방식, 즉 젊은 층에는 채권보다 주식의 편입비율을 더 높이고 반대로 고령층에는 주식보다 채권의 편입 비율을 더 높이는 방식을 자동으로 하거나, 아니면 고객이 직접 자신이 원하는 주식과 채권 비율을 입력하면 거기에 맞춰 최적의 포트폴리오를 구성하는 방식으로 구현할 수 있다.

최근에는 인공지능 학술분야에서도 로보 어드바이저의 알고리즘과 효과에 대해 다양한 논문들이 나오고 있으며, 전통적인 금융수학적인 포트폴리오 이론을 충실히 따르는 것부터, 강화학습을 이용한 에이전트 기반의 알고리즘을 이용하는 것까지 매우 다양한 방식이 연구되고 있다.

## 5-5 레그테크와 셉테크

레그테크와 셉테크는 법률 관련 과제를 인공지능 기술로 자동화 하는 리걸테크(Legal Tech)와는 구분된다. 레그테크는 규정을 의미하는 레귤레이션(Regulation)과 기술(Technology)을 합친 신조어이며, 셉테크는 감독기관의 감독을 의미하는 수퍼비전(Supervision)과 Technology를 합친 신조어이다.

레그테크는 특히 금융기관에서 주목받고 있는데, 그 이유는 금융관련 규정을 어겼을 경우 그 벌금이 가히 천문학적이기 때문이다. 2014년 프랑스의 BNP 파리바 은행은, 미국의 이란제재를 어긴 것을 이유로 약 10조원에 달하는 벌금을 미국으로부터 부과받았고 현재 10년째 매년 1조원씩 분할 납부하고 있다. 이듬해 스탠다드차터드은행 역시 유사한 사유로 미국에서 현재까지 누적 총 1조원에 가까운 벌금을 부과받았으며, 우리나라의 기업은행도 역시 유사한 사유로 인해 1,000억원에 가까운 벌금을 미국으로부터 부과받는 등 특히 자금세탁이나 공중협박자금 조달, 미국의 대외 경제 제재 관련 금융 규정을 위반했을 경우의 벌금은 한 기관의 존립을 위협할 만큼 그 액수가 상상을 초월한다.

이처럼 금융사에서는 자사가 각종 규정을 잘 준수하고 있는지를 상시 감시해야 할 필요성이 매우 높으며, 이 모든 감시 작업을 수작업으로 진행하기에는 이제 그 규정의 복잡도와 감시의 대상이 너무 폭넓어져 이에 대한 자동화의 요구 및 필요성이 점차 늘어나고 있다.

특히 상품 규정 또한 까다로워 감독기관의 심사를 통과하기 위해서는 신 상품을 개발하는 과정에서 각종 상품관련 규정을 잘 지키고 있는지 확인하는 절차를 고도로 효율화해야 할 필요가 있으며, 이러한 모든 과정을 자동화 하려는 시도를 바로 레그테크(RegTech)라고 부르는 것이다.

한편, 각종 규정의 등장으로 인해 금융기관을 관리 감독하는 감독기관의 입장에서도 업무가 복잡해지고 다양해지는 것은 당연하다. 감독기관에서는 새롭게 등장하는 상품의 약관 등을 심사할 때 여러 금융기관에서 신고되는 수 많은 상품을 일일이 수작업으로 확인하는 것은 시간과 비용 측면에서 그다지 좋은 선택이 아닐 수 있다. 이 때문에 이러한 일련의 감독관련 업무를 자동화하려는 시도 역시 활발히 일어나고 있는데 이러한 것을 바로 셉테크(SupTech)라고 부르는 것이다.

## 5-6 RPA

RPA는 Robotic Process Automation의 약자로서, 우리 말로는 “로봇식 공정 자동화 과정” 정도로 번역할 수 있다. 그 핵심은 일상적인 비즈니스 프로세스를 자동화 하여 혁신을 얻는 과정으로 요약할 수 있다.

그러나, 여기에서의 로봇 역시 팔다리가 달린 실물 로봇을 의미하는 것이 아닌 챗봇의 예처럼 소프트웨어 로봇<sup>18)</sup>이나 인공지능 기술 등을 의미한다.

RPA를 도입하는 가장 큰 목적은 업무 처리 속도를 상당히 개선하면서 동시에 오류 발생률을 획기적으로 낮추자는 것이다.

---

18) 이를 소프트웨어 로봇은 통상 봇(Bot)이라 부른다. 챗봇이 대표적인 예이다.



금융권에서도 업무 자동화를 위한 RPA 도입이 활발히 이루어지고 있는데, 주로 단순 반복하여 일어나는 시간소비형 작업을 자동화함으로써 큰 업무 절감을 꾀하고 있다. 주된 단순 작업으로는 데이터의 수집 업무, 데이터의 입력과 관련된 업무, 단순 비교업무 등이 있으며 이와 연계된 재무, 회계, 고객관리 등의 과제에 시간과 비용을 단축하면서도 동시에 정확성을 기할 수 있는 것이다.

몇 가지 RPA 사례로는 보험사의 소액 보험금 자동지급 시스템, 보험 청약서류의 자동 심사 시스템, 은행의 대출 신청 서류의 대출자격 적격성 심사 시스템 등이 있으며, 여러 다양한 분야에 있어 RPA 도입이 가능하다.

최근에는 인공지능 기법이 접목되면서 보다 복잡하고 다양한 영역으로까지 RPA 도입을 확대하고 있는 추세이다.

## 5-7 챗봇

챗봇은 잡담을 의미하는 영문 채팅(Chatting)과 로봇(Robot)의 합성어이다. 이때 로봇 역시 RPA에서처럼 물리적인 형태를 가진 것을 의미하는 것이 아닌 소프트웨어의 가상 자동화 도구를 일컫는 용어인 봇(bot)을 합성한 것이다.

챗봇의 핵심은 ‘자연어 처리’라는 인공지능 기술이다. 자연어란 인공어의 반대 개념으로서 사람이 인공적으로 만든 컴퓨터 언어 등과 달리 자연 발생적으로 생긴 언어로서 인류가 일상적으로 사용하고 있는 한국어, 영어 등의 모든 언어를 일컫는 말이다.

챗봇은 크게 3가지 유형으로 나눌 수 있는데, 가장 하위 단계이며 초보적인 단계는 간단한 키워드 정도 위주로 파악하는 기능과 함께 미리 정해진 질문 몇 가지만을 집중적으로 이해하는 기초단계이다. 이때 미리 정해진 질문을 ‘폐쇄형’ 질문이라고 한다. 한편 미리 정해진 형식이 없이 모든 유형의 질문은 ‘개방형’ 질문이라고 한다.

챗봇의 두 번째 단계는 이러한 개방형 질문을 주로 이해할 수 있는 발전된 단계로서, 다양한 디지털 비서들 즉 애플의 시리, 삼성의 빅스비, 구글의 구글 어시스턴트 등이 그 대표적 사례이다.

챗봇의 마지막 단계는 전체 맥락과 문맥 그리고 대화하는 사람의 감정까지 이해할 수 있는 단계로서 자연스러운 조언까지 가능한 단계이다. 현재 금융기관에서 상용화되어 있는 챗봇은 고도의 대화를 이어갈 수 있는 최첨단 기술을 적용한 것이 아니라 주로 키워드 위주로만 파악하고 폐쇄형 질문 몇 가지로 운영되는 1단계인데, 이는 기술이 부족해서 생긴 문제라기보다는 고객들이 주로 던지는 질문의 유형과 그를 상대하기 위한 비용 효율적 측면에서의 절충으로 볼 수 있다.

향후 자연어처리의 비약적 발전과 더불어 이러한 챗봇의 활동범위는 보다 확대될 것으로 예상된다.

앞서 인공지능의 정의를 개략적으로 설명할 때, 머신러닝이란 인공지능의 한 분야로서 데이터 사이에 숨겨진 패턴을 기계 스스로 찾아내도록 하는 알고리즘을 연구하는 학문분야라고 소개했다. 또, 딥러닝은 이러한 머신러닝의 일종으로서 심층 신경망이라는 특수한 아키텍처를 사용하는 기법이라는 것도 살펴보았다. 이 절에서는 이 각각의 기법에 대해 보다 자세히 살펴보도록 하자.

## 1 머신러닝의 이해

머신러닝(Machine Learning)이라는 용어는 1959년 미국의 아서 사무엘(Arthur Samuel)이라는 IBM의 공학자가 최초로 사용한 것으로 알려져 있다. 아서 사무엘은 머신러닝을 다음과 같이 정의했다.

“머신러닝이란 명시적으로 프로그래밍하지 않고도 컴퓨터가 스스로 학습할 수 있는 능력을 부여하는 방법을 연구하는 학문이다.”

위 설명에서 강조하고 있는 부분은 사람이 처음부터 마지막 결과에까지 이르는 모든 과정을 명시적으로 프로그램 하여 정해진 방향을 일일이 지시하는 기존의 프로그램 방식이 아니라는 점이다. 즉, 처음부터 최종 결과에 이를 때까지 컴퓨터가 데이터를 사용해 ‘스스로’ 학습해 가며 최적의 해를 찾아나가는 것이며, 이를 연구하는 모든 학문은 머신러닝이라 부른다는 것이다.

한편, 머신러닝의 보다 정교하고 실질적인 정의를 내린 사람은 미국의 전산학자 톰 미첼(Tom M. Mitchell)이다. 그는 머신러닝 기법을 다음과 같이 형식적으로 정의했다.

“컴퓨터 프로그램이 특정 과제 부류 T와 성능 척도 P에 대해 경험 E로부터 학습했다는 의미는, P로 측정된 과제 T에서의 성능이 경험 E에 대해 개선되었다는 것이다.”

마치 동일한 언어가 반복되는 듯한 상당히 복잡해 보이는 이 문장의 의미는, 다음과 같이 몇 가지로 분해하여 하나씩 나눠서 생각해 보면 좀 더 쉽게 이해할 수 있다.

- 1) 학습의 주체는 컴퓨터 프로그램이다.
- 2) 컴퓨터 프로그램의 성능 변화는 P라는 척도를 사용해 측정한다.
- 3) 컴퓨터 프로그램은 E라는 경험을 바탕으로 학습한다.

E라는 경험은 현재 손에 쥐고 있는 과거의 데이터를 의미한다. 머신러닝이란 이 과거 샘플 데이터에 기반하여 스스로 학습하는 방법에 관한 것인데, 기계가 보다 나은 방향을 향해 학습되고 있는지 판단할 수 있는 기준인 P가 존재하여, 이 P라는 척도로 현재 상태를 측정했을 때, 그 성능이 개선된 것으로 나타난다면 비로소 ‘학습이 되었다’고 판단할 수 있다는 의미이다. 이는 앞서 언급한 지도학습의 설명과 그대로 일치한다.

머신러닝이 기존의 프로그래밍과 가장 다른 점은 컴퓨터 스스로 규칙을 찾아내도록 한다는 것인데, 이는 달리 말하면 사람의 개입이 줄어든다는 의미가 된다. 사람의 개입이 줄어든다는 의미는 보다 더 많은 정보를 데이터에서 직접 얻어낸다는 의미이기도 하다. 사람이 더 많이 개입될수록, 정작 데이터에 숨겨져 있는 진정한 규칙의 정보를 추출하는 작업이 사람의 의도에 의해 방해받을 수 있다.

이 때문에 가급적 개입을 억제해 데이터 자체에서 정보를 찾아내려는 시도로 이해할 수 있고, 그 극단적인 방법이 바로 다음에 살펴볼 딥러닝이다. 우선 딥러닝으로 주제를 바꾸기 전에 머신러닝이 가지고 있는 영원한 숙제인 분산과 편향 사이의 균형에 대해 잠시 살펴보도록 하자.

## 1-1 분산과 편향

다른 모든 수학적 추정방식과 마찬가지로 머신러닝으로 예측을 수행하면 항상 오차가 발생하기 마련이다. 이렇게 발생하는 오차는 크게 분산과 편향으로 설명할 수 있는데 분산과 편향은 서로 트레이드 오프 관계에 있다.

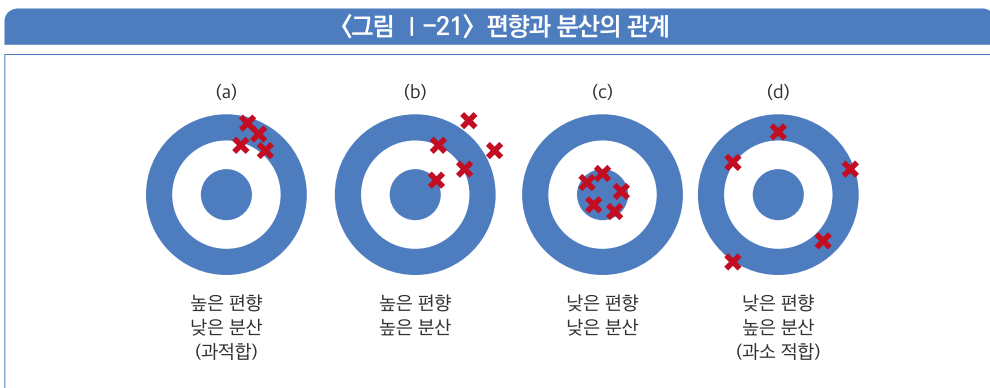
### 가. 분산

분산이란 주어진 데이터 포인트에 대한 모델 예측값의 변화 정도를 나타낸다. 모델의 분산이 높다는 의미는 훈련 데이터에 너무 많이 의존되어 일반화되기 힘들다는 의미이기도 하다. 그 결과 이러한 모델은 훈련 데이터에서의 성과는 매우 좋게 보이지만 테스트 데이터에서는 높은 오류를 일으키는 경향을 보인다.

### 나. 편향

편향은 모델의 평균 예측값과 정답 사이의 차이 분포를 의미한다. 모델의 편향이 높다는 의미는 훈련 데이터를 제대로 이용하지 못했다는 의미이다. 편향이 높은 모델은 훈련과 테스트 데이터 모두에서 높은 오류를 일으키는 경향을 보인다.

모델을 만들 때는 항상 분산과 편향 사이에서 균형을 잡기 위해 노력해야 한다. 모델과 분산에 대한 이해는 다음과 같은 과녁 그림에 비유하면 보다 쉽게 이해할 수 있다. 다음 그림을 보자.



〈그림 1-21〉은 모델의 편향과 분산을 과녁에 비유해서 그린 것이다. 그림을 보는 방법은 다음과 같다. 그림의 중심 과녁은 우리가 찾고자 하는 정답이다. 그리고 붉은색으로 표시된 것은 훈련데이터를 사용해 모델을 학습시킨 다음 예측을 수행한 결과이다.

그림의 좌상단에 있는 (a)의 경우 모두 과녁 중심에서 우상단으로만 몰려서 상당히 높은 편향을 보여주고 있다. 그러나, 그 평균적인 예측값은 모두 유사한 형태를 보여주고 있으므로, 분산은 낮다. 그러나 이 예측값은 모두 과녁 중심의 정답과는 거리가 멀어, 일반화되지 못했음을 보여준다.

이는 주어진 훈련 데이터에만 너무 충실히 의존했기 때문에 훈련 데이터에는 좋은 성능을 보이지만 테스트 데이터에서는 좋은 성능을 보이지 못하는 결과를 초래한다. 이런 경우를 일반적으로 과적합이라고 부른다. 훈련 데이터에만 너무 과하게 적합화 되었음을 일컫는 말이다.

그림 우 상단에 있는 b)의 경우도 모두 과녁의 우측에 예측이 쏠려 있어 한쪽으로 매우 편향되어 있음을 알 수 있다. 또한 예측값 역시 우 상단에 퍼져있어 분산 또한 높음을 알 수 있다.

그림 좌 하단의 c)는 가장 바람직한 모습으로서 예측값이 모두 과녁의 중앙에 정확히 몰려있어 편향도 낮고 동시에 분산도 낮은 모습을 보여주고 있다. 머신러닝의 궁극적인 목표는 이러한 편향도 낮고 분산도 낮은 값의 모델을 찾아내는 것이다.

그림 우 하단의 d)는 예측값이 어느 한 곳으로 몰려있지 않아 편향이 낮지만 예측값은 여러 곳에 흩어져 있어서 분산이 매우 높음을 보여준다. 이 경우는 훈련 데이터 내에 있는 규칙을 충분히 찾아내지 못한 것으로 볼 수 있으며, 일반적으로 과소적합이라고 부른다. 이는 훈련 데이터에 제대로 적합화되지 못했다는 의미가 된다.

## 2 딥러닝의 이해

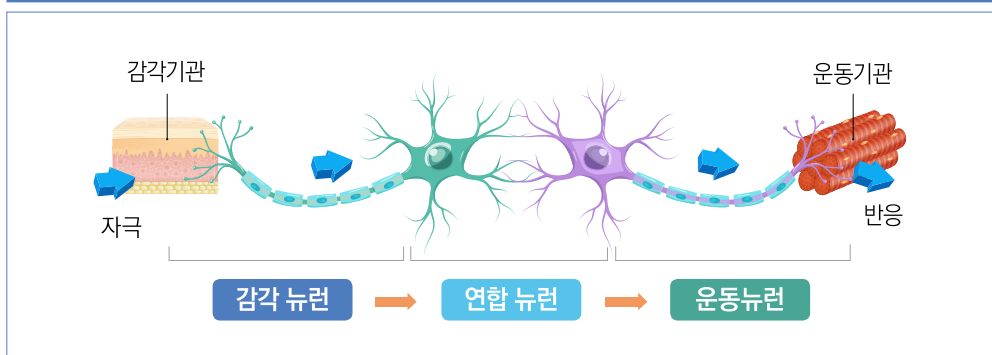
앞서 몇 차례 언급한 것처럼 딥러닝은 머신러닝의 일종이다. 머신러닝의 정의가 컴퓨터 스스로 학습하는 알고리즘에 대한 연구에 관한 것이었으므로, 딥러닝 역시 사람의 개입이 최소화되고 컴퓨터 스스로 데이터에서 정보를 찾으려 학습하는 알고리즘을 연구하는 분야로 생각할 수 있다.

### 2-1 퍼셉트론의 등장

딥러닝의 기초가 된 것은 1957년 발명된 프랭크 로젠블랫(Frank Rosenblatt)의 퍼셉트론(perceptron)이다. 뇌 과학자들은, 인간은 신경세포로도 불리는 뉴런이라고 불리는 신경계의 감각기관이 시냅스라 불리는 연결체들로 복잡하게 얽혀 있으며, 이 연결체들은 서로 전기신호를 주고받으며 외부의 자극을 뇌를 비롯한 각종 기관에 전달한다는 사실을 알아냈다.

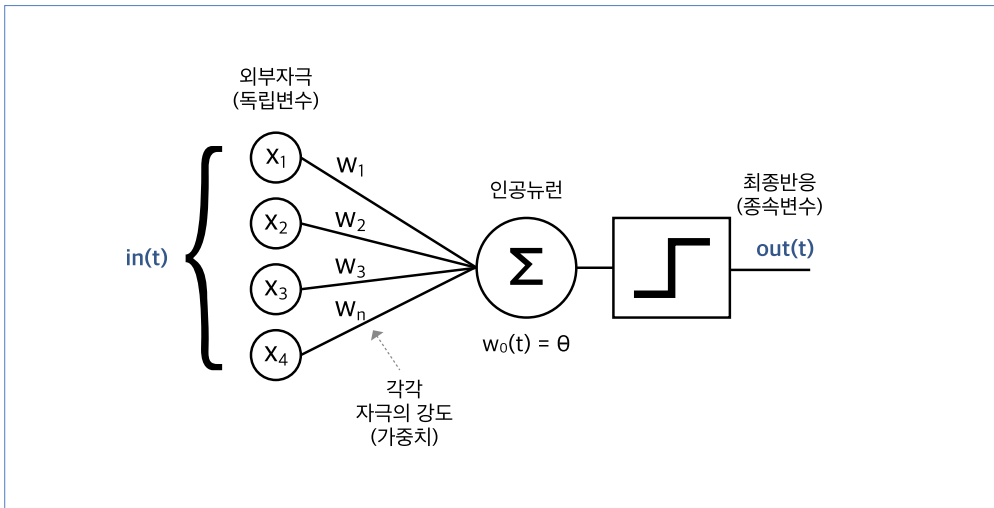
이때 외부의 자극은 전류로 바뀌는데, 자극이 매우 적을 때는 임계치를 넘지 못해 하나의 뉴런에서 다른 뉴런까지로 이 자극이 적절히 전달되지 않고, 어느 일정 부분 이상의 자극이 들어올 때만 시냅스를 통해 하나의 뉴런에서 또 다른 뉴런으로 자극이 전달되어 감각기관으로부터 운동기관으로까지 전달된다는 사실을 알게 되었다.

〈그림 1-22〉 뉴런과 시냅스



로젠블랫은 이러한 뉴런과 시냅스 관계를 전자 회로를 통해 구현한다면 인간의 뇌 기능을 일부 기계적으로 흉내 낼 수 있을 것이라 생각했고 이에 따라 인간의 뇌를 그대로 전자적으로 구현하려는 시도를 하였으며, 이를 퍼셉트론(Perceptron)이라는 이름으로 실제로 구현하였다.

〈그림 1-23〉 퍼셉트론의 기본 개념



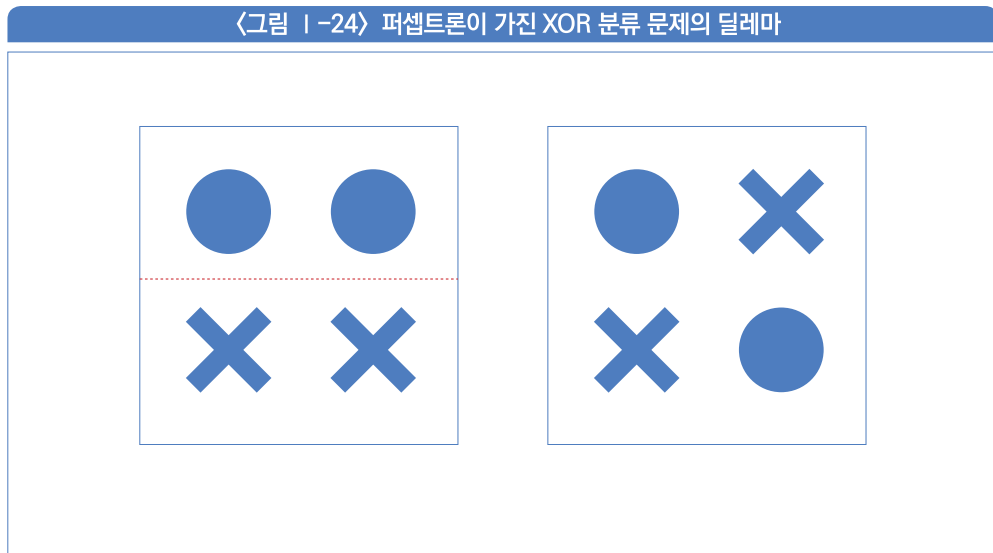
〈그림 1-23〉는 퍼셉트론의 개념도를 보여준다. 그림에서  $x_1 \sim x_n$ 은 외부의 자극이며, 수학적으로는 독립변수에 해당한다. 이 외부자극은 각각의 가중치, 즉 자극 강도( $w$ )를 가지고 모두 인공 뉴런으로 전달된다. 인공 뉴런에 전달된 모든 자극은 합산되어 최종 출력으로 보내지는데, 이 최종 출력은 계단함수를 거치면서 일정 이상의 값일 경우는 1로 반응하고, 그 이하에는 0 값을 출력하며 반응하지 않도록 설계돼 있다.

프랭크 로젠블랫은 이러한 전자 회로를 충분히 복잡하게 만들면 인간의 뇌를 흉내낼 수 있을 것이라 믿었다.



## 2-2 퍼셉트론의 한계

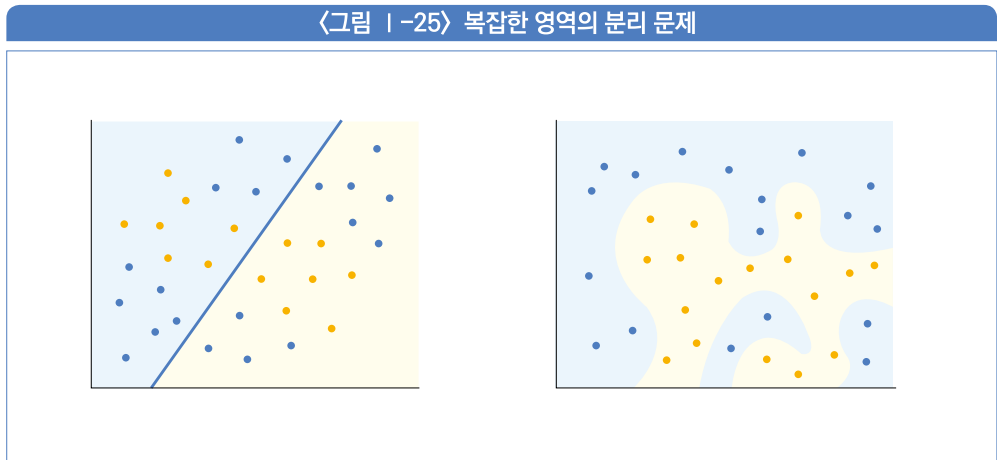
그러나 퍼셉트론은 근본적으로 한계를 가지고 있었다. 간단한 분류문제도 해결하지 못하는 퍼셉트론의 한계를 이해하기 위해 다음 그림을 살펴보자.



〈그림 1-24〉은 직선을 하나 그려 ●와 × 영역을 완전히 분리하는 문제의 예시를 보여 주고 있다. 그림의 왼쪽을 보면, 가운데 붉은 점선을 하나 그려 쉽게 ●와 × 영역을 완전히 분리할 수 있음을 보여준다.

그러나 그림의 오른쪽을 살펴보자. 그림의 오른쪽의 경우는 어떤 방향으로 직선을 긋더라도 하나의 직선으로는 ●와 ×의 영역을 완전히 분리할 수 없음을 보여준다.

이러한 딜레마는 문제 영역이 더 복잡해지면 더 극대화되는데 예를 들어 <그림 1-25>와 같은 복잡한 영역의 분리에는 직선으로는 도저히 표현할 수 없음을 잘 알 수 있다.

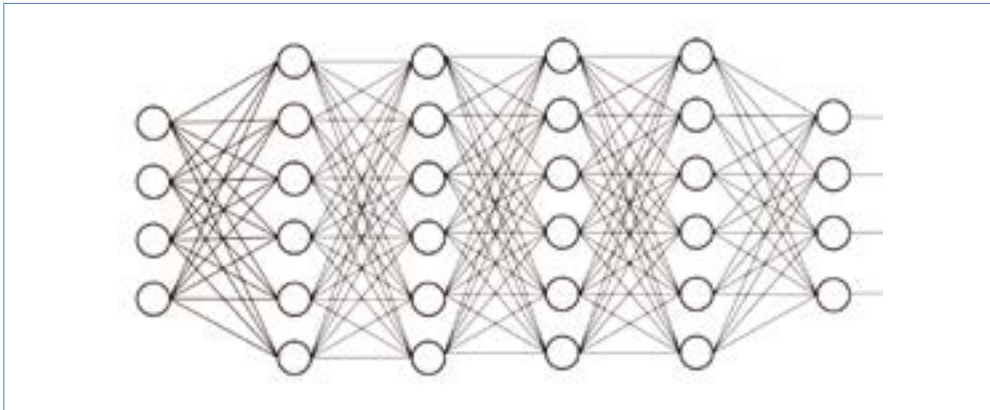


퍼셉트론의 근본적 한계는 모든 자극(독립변수)이 인공 뉴런에서 합쳐진 후 그 값이 출력할 때 모두 변수들 사이의 관계식이 선형으로만 이루어져 있기 때문에, 오로지 선형으로 된 관계만 표현 가능하다는 것이다. 그로 인해, 선형으로는 표현할 수 없는 보다 복잡한 분류 작업은 원천적으로 표현이 불가능한 근본적인 문제를 가지고 있는 것이다.

### 2-3 심층 신경망의 탄생

현대의 심층 신경망과 로젠블렛의 퍼셉트론은 크게 두 가지 측면에서 다르다. 첫째는 인공 뉴런에서 출력되는 값이 단순히 변수들 사이의 선형식이 아닌 그 출력이 비선형식으로 변환된다는 것이다. 둘째는, 인공 뉴런들이 단일 계층이 아니라 여러 겹으로 포개져 구성되어 있다는 것이다. 반드시 있어야만 하는 처음의 입력 계층과 제일 마지막의 출력 계층을 제외하고 그 사이에 존재하는 모든 계층을 보통 은닉 계층이라고 부르는데, 신경망은 여러 겹의 은닉층으로 이루어져 있다. 다음 그림을 보자.

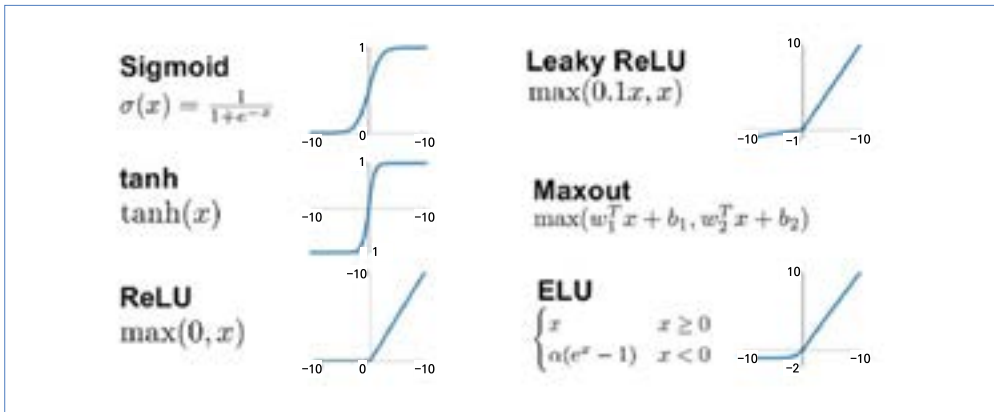
〈그림 1-26〉 여러 겹으로 쌓인 심층 신경망의 구조



〈그림 1-26〉는 인공 뉴런들이 여러 은닉층들이 겹겹이 쌓인 심층 신경망의 구조를 보여주고 있다. 복잡한 과제일수록 이 계층 수는 늘어나고, 계층 수가 늘어날수록 보다 어려운 문제를 해결할 수 있다. 2015년 이미지 인식분야에 있어 사람의 능력을 뛰어 넘었던 심층 신경망인 ResNet의 구조는 무려 152개의 계층으로 이루어져 있다.

선형을 비선형으로 바꾸는 방법은 단순히 퍼셉트론의 출력인  $y$  직전에 비선형 함수의 입력으로 주어  $y$ 를 변형하는 것이다.  $X \rightarrow y$  함수를  $X \rightarrow z \rightarrow f(z) = y$ 로 변형한 것이다. 이때 변형에 관련된 함수  $f(\ )$ 를 '활성 함수(Activation function);라 부른다.

〈그림 1-27〉 활성화 함수의 예시



〈그림 1-27〉은 일반적으로 많이 사용되는 비선형 함수를 보여주고 있다. 각 함수는 응용분야에 따라 서로 장단점이 있으며, 모두 미분 가능한 특징을 갖고 있다.<sup>19)</sup>

## 2-4 범용 근사 정리

신경망의 수학 이론에서 범용 근사 이론(Universal Approximation Theorem)은 연속 함수를 사용한 신경망으로 구성된 심층 신경망 구조는 한정된 수의 뉴런 개수를 가진 하나의 은닉층만으로 이루어지더라도 모든 연속 함수를 표현할 수 있다는 것을 증명하고 있다. 즉, 이론적으로 심층 신경망을 사용하면, 생각할 수 있는 아무리 복잡한 결정 경계라도 다 표현할 수 있다는 의미가 되며, 기본적으로 존재하는 모든 분류 문제를 해결할 수 있다는 의미가 된다.

## 2-5 심층 신경망의 역할과 한계

딥러닝 기술은 특히 현대의 빅데이터와 함께 그 효용이 극대화되고 있다. 전통적인 머신러닝 기법과 달리, 딥러닝 기법은 모든 입력을 단순히 인공 뉴런에 전달하기만 하면, 최적의

19) ReLU의 경우  $x=0$  일 때, 0으로 정의해서 미분 가능하게 처리

모델 가중치를 컴퓨터가 스스로 학습하기 때문에 특히 비정형 데이터를 다루기에 적합하다.

이 때문에 음성이나 이미지, 영상 처리 등은 예외 없이 딥러닝 기법을 사용하며, 번역 등의 자연어 처리 영역에서도 딥러닝을 사용하기 시작하면서 비약적으로 성능이 향상되고 있다.

그러나, 딥러닝은 사람의 개입이 최소화 되므로 소위 ‘블랙박스’ 기법이라는 근본적인 한계 또한 가지고 있다. 즉, 어떤 결과가 나왔을 때, 왜 그런 결과가 나왔는지에 대한 논리적 해석은 거의 불가능하다. 그저 결과를 얻을 뿐 그 이유를 알 수 있는 어떤 논리적 가정 등은 개입되지 않기 때문이다.

만약 딥러닝으로 얻은 결과가 일반 상식에 반한다면, 그 결과를 단순히 믿어야 할 것인지에 대한 결정의 문제에 직면하게 된다. 이러한 문제는 제7절 5의 ‘해석 가능한 인공지능’에서 다시 한번 살펴보기로 하자.

### 3 금융과 인공지능

금융 분야에서도 인공지능은 점점 보편적으로 사용되고 있다. 금융에서 일반적으로 사용되는 여러 비즈니스 사례는 제5절 5에서 이미 자세히 살펴보았다. 금융데이터는 몇 가지 특성을 가지고 있는데 그중 가장 두드러진 특징은 시간에 연계되어 발생하는 시계열 데이터라는 것이다.

시계열(TimeSeries) 데이터란 특정 시간에 대해 반복적으로 측정되어 얻는 관측치를 의미한다. 이러한 데이터를 시각화 한다면 한 축(대체로 가로축)은 통상적으로 시간 축이 된다.

〈그림 1-28〉 주가 그래프



〈그림 1-28〉은 대표적인 시계열 데이터 중 하나인 주가 그래프를 보여주고 있다. 그림에서의 가로축은 시간이며, 세로축은 주식의 가격이다. 이 그래프는 시간의 흐름에 따라 주가가 변동하는 모습을 보여주는 대표적인 시계열 데이터이다. 시계열 데이터는 동일한 주기에 대해 데이터를 수집하는 경우와 일정하지 않은 주기 간격으로 데이터를 수집하는 두 가지 종류가 있다.

이러한 시계열 데이터는 다음과 같은 몇 가지 특성으로 인해 조심해서 다루어야 한다.

### 3-1 불변성(immutability)

시계열 데이터는 시간 순서에 따른 관측이므로 항상 새로운 개체가 생성되어 추가된다. 따라서 이미 수집된 데이터는 '변경 불가'이어야 하며 항상 추가만 가능(append-only)해야 한다.

### 3-2 자기상관(Auto Correlation)

시계열 데이터의 특정 데이터는 그 데이터 이전과 이후의 상태와 밀접한 관련을 지니는 경우가 많다. 예를 들어 <그림 1-28>의 주가 데이터 역시 오늘 날짜의 주식가격은 당연히 어제 날짜의 주식가격과 완전히 밀접된 관계를 가지고 있다. 또한 내일의 주식가격에 밀접한 관계를 가진다. 자기상관이 있다면 특정 시각  $t$ 와 그의 지연된 시각  $t-i$  사이에 선형의 관계가 형성된다. 자기상관을 가진 데이터는 회귀분석 등 일반적 분석을 방해하거나 불가능하게 한다.

데이터에서 자기상관을 제거하는 다양한 방법이 나와 있고 이를 적절히 활용할 필요가 있다. 자기상관은 계열 상관(Serial Correlation)이라고도 부른다.

### 3-3 계절성(Seasonality)

금융 데이터의 시계열 데이터이기 때문에 발생하는 또 다른 특징은 계절성이다.

계절성이란 데이터에서의 어떤 변동성이 특정한 주기로 나타나는데 그 주기가 1년 미만일 경우를 칭한다. 예를 들어 매년 7월~8월 휴가철에 나타나는 월별 변동성이 유사하거나 혹은 어떤 특정 이벤트로 인해 일별, 분기별, 월말별 등 특정 주기를 따라 유사한 패턴이 반복되는 것을 의미한다.

필요할 경우, 금융 데이터에서 주요 정보를 추출하기 전에 계절성이 미치는 영향을 제거할 수 있는 다양한 기법을 적용할 필요가 있다.

### 3-4 그 밖의 특성

그 밖에 금융데이터 그중에서도 주식 수익률과 같은 자산 수익률 데이터에는 여러 가지 특성이 있다. 대표적인 것이 일별 수익률의 로그값은 정규분포를 따른다는 가정이다.

금융 데이터를 사용해 인공지능 기법을 활용할 때는 이러한 기본 가정을 통해 특정 패턴의 반복성을 알아내고자 시도하기도 한다.

최근에는 주식 시세 예측과 관련되어 강화학습을 응용하는 등 단순한 데이터의 패턴 분석에서 벗어난 다양한 시도가 이루어지고 있다.



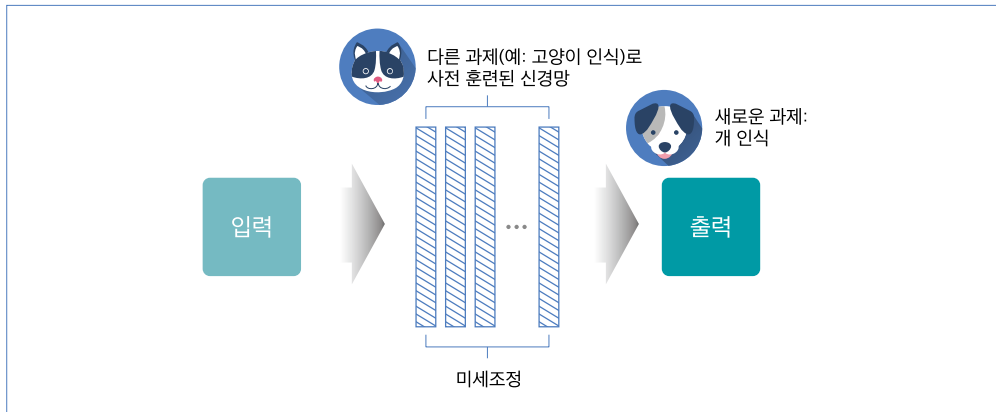
지금까지 인공지능과 빅데이터의 정의와 기본 작동원리를 매우 간략히 살펴보았다. 제7절에서는 인공 지능 기술이 비약적으로 발달한 배경이 된 몇 가지 기본 개념과 함께 인공지능이 할 수 있는 일과 할 수 없는 일을 구분해 보도록 한다. 인공지능으로 할 수 있는 일과 할 수 없는 일을 구분하는 것은 현업에서 인공지능을 도입할 때 주요한 참고가 될 수 있을 것이다.

## 1 전이학습

인공지능으로 모델을 구축한다는 것은 각 입력변수에 할당할 적절한 가중치를 찾아내는 것이라 할 수 있다. 통상적인 절차는 최초의 가중치를 임의로 할당하거나, 몇 가지 규칙에 의해 할당한 다음 최적의 값을 찾을 때까지 반복 학습해 나가는 과정을 거친다. 이때 최초의 가중치는 찾고자 하는 최적의 가중치와 매우 다를 것이므로 최적값을 찾아나가는 과정은 대체로 시간이 오래 걸리고 계산 자원이 많이 소요될 수밖에 없다.

전이학습(Transfer Learning)은 이러한 단점을 극복하고자 하는 시도에서 비롯되었다. 예를 들어 이미지가 입력으로 주어지면 고양이인지 아닌지 찾아내는 인공지능 모델을 성공적으로 훈련시켰다고 가정해 보자. 이 모델은 아주 오랜 훈련을 거쳐 주어진 이미지가 고양이가 맞는지 알아낼 수 있는 특정 가중치 값을 알아냈을 것이다. 어렵게 알아낸 이 가중치를 재활용할 수 있는 방법이 있을까?

〈그림 1-29〉 전이학습의 개념도



〈그림 1-29〉는 전이학습의 개념도를 보여준다. 그림은, 고양이를 인식하도록 훈련된 모델을 가지고 와서 개를 인식하는 새로운 과제에 응용하는 예제를 보여준다.

이 새로운 과제를 위해 완전한 랜덤 가중치를 부여하여 다시 처음부터 찾아나가는 대신, 고양이에 훈련된 가중치를 가져와 그를 시작점으로 해서 그 가중치를 ‘보장’해 나가는 방법을 활용하는 것이다.

이런 식으로 훈련하면 완전히 처음부터 훈련하는 것에 비해 시간과 계산 자원을 크게 절약할 수 있고 성능 또한 크게 개선할 수 있다.

신경망 학습에 있어서의 전이학습이라는 개념은 스티보 보치노프스키(Stevó Bozinovsk)와 안트 풀고시(Ante Fulgosi)가 1976년에 발표한 논문에서 비롯되었다.

전이학습은 머신러닝의 급격한 발전에 기여했으며, 매우 보편적으로 사용되고 있는 기술 중 하나이다.

## 2 연합학습

연합학습(Federated Learning 또는 Collaborative learning)은 소위 에지(Edge) 디바이스들이 연합해 각각의 로컬에서 훈련이 일어나는 방식이다.

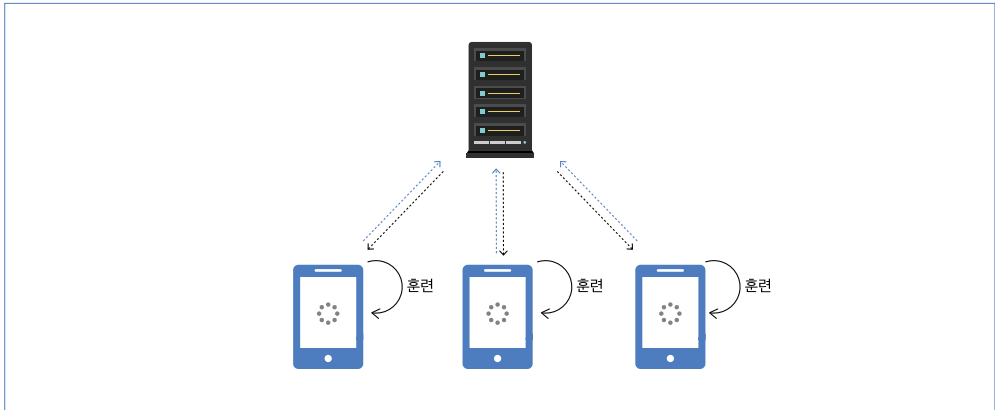
### TIP

예전에는 에지(Edge) 디바이스가 통신장비 그중에서도 주로 라우터를 의미하는 것이었지만, 현재는 통신 기능을 갖춘 사물 인터넷 기기를 통칭하는 것으로 사용된다. 특히 스마트 폰 등 통신기능과 연산 기능을 모두 갖추고 있는 휴대용 단말기 등도 모두 에지 디바이스라고 부른다. 에지 디바이스는 통신에 관여할 수 있는 다양한 형태로 존재한다.

전통적인 인공지능 훈련 방식은 모든 데이터를 중앙 서버에 업로드하면서 시작된다. 한 곳에 모든 데이터가 모이면 이 데이터를 사용해 모델을 훈련하는 것이 일반적인 방식이다.

그러나 연합훈련에서는 모든 훈련이 에지 디바이스에서 일어나고 중앙 서버는 그 결과 가중치 값만 전달받는다. 따라서 에지 디바이스에 있는 데이터가 개인정보 등 프라이버시에 관련된 데이터라 할지라도 중앙 서버에 전달하지 않아도 되므로 보안문제에 대한 걱정을 하지 않아도 되는 장점이 있다. 또한 분산된 상태에서 각자의 계산 자원을 사용하므로, 중앙 서버가 막대한 계산 자원을 절약할 수 있는 장점도 동시에 얻을 수 있다.

〈그림 1-30〉 연합훈련 방식



〈그림 1-30〉는 전형적인 연합훈련 방식을 보여주고 있다. 그림에서 보는 것처럼 실제 훈련은 각각의 에지 디바이스에서 일어나므로, 데이터의 직접적인 교류가 없으므로 데이터 전송을 위한 시간 지연이 없음은 물론 민감한 데이터의 전송으로 인한 보안의 문제도 사라진다. 특히 최근의 첨단 스마트 폰들의 계산 성능은 웬만한 데스크톱보다 오히려 더 우수할 정도로 에지 디바이스의 성능은 뛰어나다.

### 3 자동 머신러닝(Automated ML, AutoML)

과거 컴퓨터 프로그래밍은 주로 전산학과에서 전문적 교육을 받은 소수의 사람들만 다룰 수 있을 정도로 쉽지 않은 영역이었다. 컴퓨터 프로그램은 컴퓨터라는 기계와 소통하기 위해 사람들이 인위적으로 만든 프로그램 언어라는 것을 통해 명령을 내린다. 과거에는 이 프로그램 언어를 습득하기 위해서는 하드웨어의 기본 작동원리까지 알아야 하던 시절도 있었지만, 현대는 다양하고 편리한 언어들이 대거 쏟아지면서 이러한 제약은 이제 옛말이 되었다.

최근에 각광받고 있는 파이썬이나 R 등의 언어는 전산학의 기본지식 없이도 얼마든지 쉽고

간편하게 습득할 수 있으며, 또한 이 언어의 능력을 배가시켜줄 수 있는 기능들의 집합인 라이브러리가 풍부하게 제공되고 있다.

이제 인공지능을 하기 위해서는 복잡한 전산 교육 없이도 파이썬 등의 간단한 언어와 함께 파이썬에서 제공하는 다양한 인공지능 라이브러리를 간단히 익히기만 하면 누구나 쉽게 인공지능을 시작할 수 있을 정도로 간편해졌다.

편의성을 위한 사람들의 욕구는 여기서 그치지 않고, 보다 더 편리한 방법으로 인공지능을 이용할 수 있는 방법을 연구하게 되었고, 그 중심에는 파이썬 같은 컴퓨터 언어 자체를 사용할 필요 없이 마우스 클릭 몇 번만으로 이용할 수 있는 방법도 연구하게 되었다.

자동 머신러닝은 그중의 하나로서 마우스 클릭만으로 보편적인 인공지능 기법을 모두 사용할 수 있게 해주려는 시도이다.

자동 머신러닝은 줄여 오토 ML(Auto ML)로도 부른다. 오토 ML은 머신러닝을 실생활에 적용하려는 일련의 과제를 자동화하는 프로세스를 연구하는 분야이다. 오토 ML의 궁극적인 목적은 인공지능에 대한 어떠한 지식도 없는 완전한 비전문가라도 인공지능 기법을 활용할 수 있게 해 주는 데 있다.

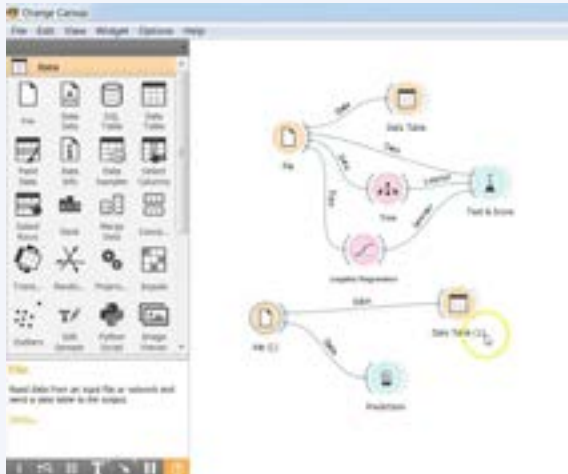


〈그림 1-31〉는 구글 클라우드를 통해서 제공하고 있는 Auto ML의 화면을 캡처한 모습을 보여주고 있다. Auto ML을 사용하면, 사용자가 해야 할 일은 데이터를 준비하는 것이 전부이다. 사용자가 데이터를 입력만 해주면 별도의 설정을 하지 않더라도 기본 설정에 의해 적절한 알고리즘을 스스로 찾아서 데이터 속에 숨겨진 패턴을 찾기 위한 훈련을 시작한 다음, 그 결과를 출력해 준다. 이 과정에서 사용자가 알아야 하거나 배워야 할 별도의 인공지능 기술은 없다. 기술을 익힌다면, 약간의 맞춤형 설정을 통해 더 나은 성능을 위해 직접 조작할 수도 있지만 그렇지 않더라도 기본 설정만으로도 계산을 훌륭하게 해 낸다.

## MEMO

이 밖에도 머신러닝을 쉽게 이용할 수 있도록 배려한 여러 시각적 도구들이 존재한다. 대표적인 것이 오렌지(orange)라는 프로그램이다.

오렌지는 1996년에 배포된 무료 소프트웨어로서 머신러닝 관련된 대부분의 기능을 마우스 클릭 정도로 쉽게 이용할 수 있도록 만들어진 소프트웨어이다.



오렌지는 별도의 전문 지식 없이도 누구나 쉽게 인공지능을 접할 수 있도록 도움을 주는 시각적 도구이다.

#### 4 설명 가능 인공지능(XAI, Explainable AI)

인공지능 앞에 굳이 설명 가능(Explainable)이라는 수식어가 필요한 이유는 일반적으로 인공지능 기술은 설명 가능하지 않다는 점을 반증하고 있다.

머신러닝 기술 그중에서도 딥러닝 기술은 사람의 개입이 최소화된다. 따라서 입력 변수들 간의 관계와 그 입력 변수가 출력 변수에 어떠한 영향을 미치는 것인지에 대한 기본적인 가정과 전제가 생략된다. 이 경우, 컴퓨터가 최적해를 찾아나가는 전체 과정은

완전한 블랙박스에 해당된다. 즉, 어떠한 과정으로 어떠한 이유 때문에 그러한 결론이 도출되었는지에 대해 거의 알 수 없다.

이 경우, 일반적인 상식이나 논리와는 부합하지 않은 결과가 나온다면 당황할 수밖에 없다.

딥러닝을 이용하여 새로운 신용평가 시스템을 구축하는 경우를 생각해 보자. 이 경우 고객 프로파일을 입력하면 궁극적으로 이 고객이 대출금을 충실히 변제할 것인지에 대한 딥러닝의 결과를 얻게 되겠지만 왜 그러한 결과를 얻게 되었는지에 대해서는 알 수가 없다. 만약 고객이 자신은 왜 대출을 받을 자격이 안 되는지 물어온다면, 담당자는 딱히 해줄 말이 없다. 그냥 딥러닝의 결과가 그러할 뿐 자기도 왜 그런지는 모르기 때문이다.

이처럼, 실제 업무에서는 반드시 그 결과를 설명할 수 있어야 하는 경우가 매우 많다. 또한 중요한 결과일수록 그 결과가 설명 가능해야 하고 논리적이고 상식적이어야 한다. 어떤 중요한 결정을 내려야 하는 순간 그 결과가 상식에 부합하지 않다면, 그에 따르지 않는 것이 더 합리적일 수 있기 때문이다.

설명 가능한 인공지능을 추구하기 위해서는 여러 가지 기법들이 있으며, 대개 입력 변수들 중 전체 결과에 미치는 정도가 어느 것이 큰지 알아보는 변수 중요도(Variable Importance) 검사를 수행한다. 예를 들어, 특정 변수  $x$ 를 제거하더라도 유사한 결과가 나온다면  $x$ 라는 변수는 결과에 미치는 중요성이 떨어지는 것으로 생각할 수 있다. 이런 과정을 여러 차례 거치다 보면 결과에 궁극적으로 영향을 많이 미치는 변수 군을 파악할 수 있게 되고, 이를 통해 전체적인 스토리를 구성해 볼 수 있다.

한편, 전통적인 머신러닝 기법은 사람이 상당 부분 개입되고, 이 경우는 그 결과에 대한 논리적 설명이 가능한 경우가 많다. 따라서 딥러닝으로 어떤 결과를 얻었다면, 어느 정도 설명이 가능한 전통적인 머신러닝 방법 중 하나를 사용해 동일한 결과를 얻게 되는지 여부를 검증해 볼 수 있다. 만약 유사한 결과를 얻게 된다면 전통적인 머신러닝 방법을 통해 어느 정도 스토리를 구성해서 논리를 파악할 수 있게 될 것이다.



## 5 인공지능이 할 수 있는 것과 없는 것

지금까지 인공지능에 대해 많은 것을 알아보았다. 이제 인공지능이 할 수 있는 것과 할 수 없는 것에 대해 전반적으로 정리를 해보도록 하자.

### 5-1 인공지능은 복잡한 개념을 익힐 수 없다.

사람들은 개념을 이해할 수 있다. 개념이란, 개개의 사물로부터 일반적이면서 공통적인 성질을 뽑아내어 만든 관념이라 할 수 있다.

우리는 사람의 얼굴을 보고 눈을 인식하고 코를 인식한다. 그러나 컴퓨터에게 이미지를 주고 사람인지 고양이인지 알려줄 수 있는 방법은 없다. 아니 그 방법을 찾아낸 사람은 아직 없다. 지금까지 컴퓨터가 사람과 고양이를 구분하도록 하는 인공지능 방법은 사람이미지를 아주 큰 정수로 만들고, 고양이 그림을 아주 큰 정수로 만들어, 이 두 수 사이에 어떤 차이가 있는지 스스로 찾아내도록 학습시키는 것이 유일하다.

컴퓨터는 이 숫자가 무엇인지 모른다. 다른 두 숫자 사이에 차이가 있다는 점만 알 수 있다.

사람은 다르다. 사람은 고양이라는 생물의 생김새가 어떤지 그 ‘특성’을 배우고 이해한다. 그러면 어떠한 고양이 그림이 오더라도 그 특성이 맞는지 확인하여 고양이인지 인지한다. 따라서 단 몇 장의 샘플 사진으로도 개념을 충분히 학습시킬 수 있다. 그러나, 컴퓨터는 이러한 개념을 학습시킬 방법이 없다. 컴퓨터가 고양이를 인지할 수 있는 방법은 수만 장의 고양이 사진을 각각 정수로 만들어 이 정수에서 공통된 패턴을 스스로 학습하게 만드는 것이다. 그 다음 새로운 이미지가 오면, 그 이미지를 또 다른 정수로 만든 다음, 앞서 학습한 패턴이 이 그림에 깃들여 있는지 알아보는 것이다.

MEMO

개와 늑대를 구분하는 학습의 경우를 생각해 보자. 이때 컴퓨터를 학습시킬 이미지를 생각해 보면, 대부분의 개는 가정집에 찍힌 것으로서 그 배경은 가구가 들어 있는 방이거나 집 혹은 건물일 수 있다. 반면 늑대를 찍은 사진의 배경은 대체로 숲이나 나무 등의 외야(out field)일 것이다. 즉, 두 이미지의 배경이 매우 일관되게 서로 상이할 수 있는 것이다.

〈그림 1-32〉 늑대와 개는 배경이 서로 일관적으로 다를 수 있다.



이때 두 이미지의 배경을 통제하지 않은 상태에서 그대로 컴퓨터를 학습시키면 정작 개와 늑대와 상관없이 그 배경을 보고 개와 늑대를 구분하게 될 수 있다. 컴퓨터는 늑대와 개라는 개념이 아니라 모두 수치화 된 두 이미지에서 단순히 다른 점만 찾기 때문이다.

## 5-2 데이터에 패턴이 있어야 한다.

인공지능으로 로또 번호를 예측할 수 있을까? 얼핏 생각하면 최첨단의 딥러닝 기법을 사용하면 로또 번호를 완벽히는 아니더라도 어느 정도 예측할 수 있을 듯 하다. 실제로 로또 번호를 예측해 알려준다고 주장하면서 유료 서비스를 제공하는 업체도 적지 않다.

머신러닝 기법으로 로또 번호를 예측할 수는 없다. 로또는 완전히 랜덤이며 패턴이 없기 때문이다. 머신러닝은 주어진 데이터에서 숨겨진 패턴을 학습을 통해 찾아내는 기법이다. 따라서 원래 패턴이 없는 완전한 랜덤 데이터라면 아무리 학습을 시키더라도 아무런 패턴을

찾아낼 수 없다. 이때 두 가지 흔한 오류가 발생할 수 있는데, 첫째는 패턴이 없는데도 불구하고 머신러닝이 패턴으로 착각해 보고하는 경우이며, 다른 하나의 경우는 패턴이 실제로 존재함에도 불구하고 찾아내지 못하는 오류이다.

그렇다면 데이터에 패턴이 있을지 없을지 어떻게 알 수 있을까? 정답은 직접해 보는 수밖에 없다. 그러나 대부분은 업의 본질을 잘 이해하고 있는 현장 전문가들의 경우, 직관적으로 어떠한 패턴이 있을 것이라 짐작할 수는 있을 것이며 그러한 경우라면 시도해 보는 것이 좋다.

### 5-3 해석적 닫힌 식이 있는 경우는 머신러닝이 필요 없다.

한편, 이러한 패턴이 이미 잘 알려져 해석적 닫힌 식으로 표현 가능할 경우에는 머신러닝이 필요 없다. 바로 닫힌 식을 이용해 정답을 구하면 될 일이기 때문이다. 정답을 아는데 굳이 근사해를 구하기 위해 시간을 낭비할 필요는 없을 것이다.

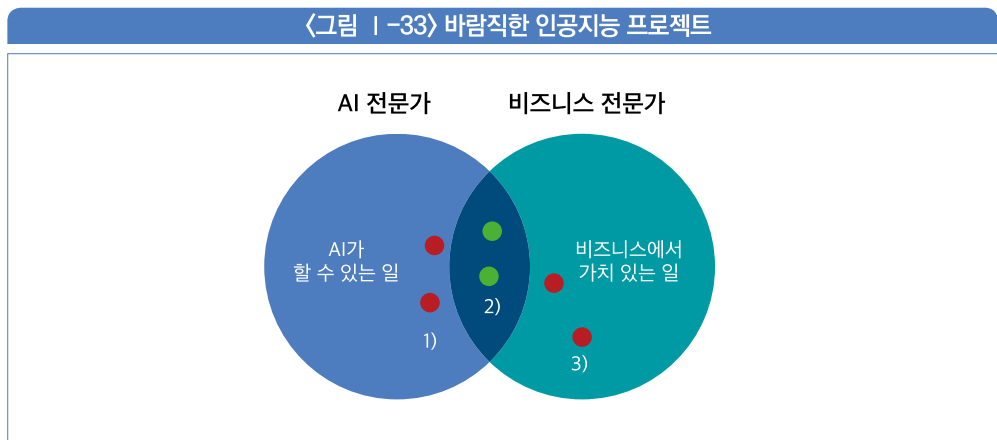
### 5-4 데이터는 무엇보다 중요하다.

인공지능의 핵심은 단연 ‘데이터’이다. 아무리 좋은 아이디어와 알고리즘이 있더라도 모델을 훈련시킬 데이터를 충분히 확보하지 못한다면 무용지물이다.

평균적 모델과 질 좋고 풍부한 데이터를 가진 경우와 최첨단 알고리즘을 가졌으나, 질이 나쁘고 양이 적은 데이터만 가진 경우를 비교해 보면 당연 질 좋고 풍부한 데이터를 가진 경우가 거의 항상 더 좋은 성능을 낼 수밖에 없는 것이다.

## 5-5 바람직한 인공지능 프로젝트

이제 끝으로 인공지능 프로젝트를 선정할 때 있어서의 유의할 점을 살펴보도록 하자. 많은 기업이 인공지능 프로젝트를 수행함에 있어 의미 있는 결과를 얻지 못하고 돈만 낭비하는 근본 원인은 비즈니스에서 가치 있는 일과 인공지능이 가능한 일 사이의 절충점을 제대로 찾지 못했기 때문이다. 다음 그림을 보자.



〈그림 1-33〉은 바람직한 인공지능 프로젝트를 보여준다. 그림의 2)번 영역은 AI가 할 수 있는 일 중에 비즈니스에서 가치 있는 일의 영역으로서 기업의 AI 프로젝트는 당연히 이 영역에서 기획되어야 한다.

그러나 많은 기업들이 그렇지 않고 오류를 범하는데 바로 다음과 같은 두 가지이다.

### 가. AI로 할 수는 있지만 비즈니스에서는 가치 없는 일

〈그림 1-33〉의 1)번 영역은 AI로 할 수는 있으나 비즈니스에는 가치 없는 일이다. 많은 기업들이 이러한 프로젝트에 돈을 낭비한다. 그 이유는 대개 두 가지인데, 첫째는 보여주기식 프로젝트를 수행하는 경우이고, 둘째는, 업의 본질을 전혀 이해하지 못한 인공지능 전문가를 외부에서 영입하고 그에게 모든 일을 맡겼을 경우이다. 특히 보여주기식 프로젝트를 추구할

경우 오로지 최첨단 기술의 구현을 좇는 우를 범하느라 거액의 돈을 낭비할 수 있다.

#### 나. 비즈니스에는 가치 있지만 시로 할 수 없는 일

사내 AI 전문가 혹은 외부에 자문할 수 있는 기술 전문가가 없는 경우에는 시로 구현하는 것이 불가능하거나 매우 힘든 프로젝트를 선정하고 시간과 돈을 낭비하는 우를 범하는 경우가 많다. 이 또한 대표적인 돈 낭비로 이어진다.

#### 다. 바람직한 프로젝트의 선정방법

바람직한 AI 프로젝트를 선정하기 위해서는 반드시 업의 본질을 잘 아는 현업 전문가와 기술을 잘 아는 AI 전문가가 협업하여야 한다. 일반적으로 성공적인 프로젝트는 현업에서 선정한 프로젝트 중 기술 전문가의 검토를 거쳐 그 타당성이 입증된 경우이다.

헬로, 핀테크!(핀테크 기반기술)



HELLO, FINTECH!

# 2장

## 클라우드

- 제1절 왜 클라우드인가? (클라우드 소개 및 효과)
- 제2절 클라우드란? (클라우드의 이해)
- 제3절 클라우드를 사용하려면? (클라우드 도입 과정)
- 제4절 클라우드 사례 (사례를 통한 클라우드의 이해)
- 제5절 클라우드와 핀테크
- 제6절 주요 클라우드 서비스 이해
- 제7절 실제로 클라우드 사용해 보기

## 2장 클라우드



### 💡 학습목표

- ① 클라우드에 대해 이해하고 비즈니스적으로 어떤 의미가 있는지 설명할 수 있다.
- ② 핀테크에서 클라우드가 어떤 역할을 할 수 있는지 분석한다.
- ③ 핀테크 기업이 알아야 할 금융 분야의 클라우드 사용 규정에 대해 알아본다.
- ④ 클라우드 서비스를 모델 및 주체, 범위에 따라 분류할 수 있다.
- ⑤ 실제 클라우드 서비스를 사용하기 위한 과정에 대해 설명할 수 있다.
- ⑥ 성공적으로 클라우드를 도입할 수 있는 방법을 설명할 수 있다.
- ⑦ 클라우드의 기본적인 기술과 서비스 들을 열거하고 설명할 수 있다.
- ⑧ 안정적인 금융서비스를 위해 중요한 기술인 보안에 대해서 알아본다.
- ⑨ 클라우드 서비스에 실제 가입하여 사용해본다.

### 💡 학습개요

클라우드는 단지 하나의 IT 기술 트렌드가 아니다. 클라우드는 금융을 포함한 전 산업에 지대한 영향을 주고 있으며, 미래 국가의 경쟁력을 결정할 수 있는 중요한 담론 중의 하나이다. 클라우드라고 하는 IT 플랫폼을 통해 기업은 IT 서비스를 아웃소싱할 수 있을 뿐만 아니라, 핀테크와 같은 새로운 비즈니스 모델을 가진 스타트업의 성장을 촉진할 수 있는 에코시스템이다. 핀테크는 전통적인 금융산업과 새로운 IT 기술 기반 서비스의 융합이라 할 수 있으며, 클라우드를 활용하면 그 효과를 극대화할 수 있다.

이에 클라우드 서비스를 핀테크 비즈니스에 적용하기 위해서는 각각의 모델과 제공자 및 사용 범위에 따른 서비스를 서비스 분류를 이해해야 한다. 그래서 비즈니스 모델에 맞는 적절한 서비스를 활용했을 때 효과를 극대화할 수 있다. 클라우드 플랫폼을 최대한 활용하고 비즈니스 아이디어를 빠르게 구현할 때는 PaaS나 SaaS가 유리할 수 있지만, 대규모 시스템을 구축하거나 기술적인 경쟁력(해자)을 가져야



한다면 IaaS 기반 접근이 유리할 수 있다. 고객 보호를 위한 규정이 많은 금융에서는 전통적으로 Private 클라우드를 많이 사용해 왔지만, 혁신적인 핀테크 기술을 빠르게 적용하기 위해서는 Public 클라우드 또는 두 가지를 결합한 Hybrid 클라우드가 효과적일 수 있다.

또한 클라우드는 IT 서비스를 빠르고 쉽게 구현할 수 있도록 플랫폼을 제공하는 서비스이다. 대부분의 경우 클라우드를 통해서 비용을 절감하고 비즈니스 서비스를 빠르게 구현할 수 있는 장점이 있다. 하지만 비즈니스 요건이 복잡해지고 클라우드를 보다 체계적으로 사용하려면 클라우드 도입에 대한 한 단계 깊은 고민이 필요하다. 이에 클라우드 도입 과정에 알아야 할 주요 내용인 클라우드 서비스 제공자(CSP, Cloud Service Provider), 클라우드 관리 서비스 제공자(MSP, Managed Service Provider), 그리고 어떤 식으로 도입 전략을 가져가면 초기 도입부터 성숙기까지 원활하게 클라우드를 사용할 수 있을지 살펴보겠다.

또한 이 책에서는 핀테크 업체가 클라우드 서비스를 도입하여 사용할 수 있는 최신 IT 기술에 대해 설명한다. 즉 클라우드 기반의 가상 머신(Computing)이나 데이터베이스 기술 등을 설명하고, 관련하여 API 서비스 등도 살펴볼 예정이다. 다만 클라우드 도입에 있어서 가장 필수적인 보안과 관련하여 특히 클라우드 사용 시 반드시 알아야 할 전자금융감독 규정을 중심으로 살펴보겠다.

마지막으로 이 책에서 설명한 클라우드 서비스를 실제로 사용해보겠다. 방대한 클라우드를 모두 경험할 수는 없지만, 그 일부분이라도 직접 해보는 것은 기술에 대한 이해를 크게 올릴 수 있다. 여기서는 실제로 클라우드 서비스에 가입하여 클라우드 사용 준비를 하는 부분부터 컴퓨팅 자원을 생성하여 필요한 서비스를 구축할 준비, 그리고 데이터를 저장하는 데이터베이스 서비스까지 실습을 해보겠다. 여러 클라우드들이 있지만 편의상 구글 GCP와 아마존 AWS를 대표로 골라 사용해 보도록 한다. 클라우드 업체들은 사용자들이 쉽게 서비스를 체험해 볼 수 있게 일부 서비스나 특정 금액까지 무료 사용을 할 수 있게 제공하고 있다. 단, 이러한 무료 서비스는 클라우드 공급자의 재량이므로, 향후에 해당 기업의 정책에 따라 변경이 될 수 있으므로 주의하기 바란다.

## 1 클라우드란 무엇인가?

### 1-1 클라우드 서비스의 정의

클라우드를 한 마디로 쉽게 정의하기는 쉽지 않다. 우선 개념이 광대하기도 하지만, 세상의 모든 것들이 그렇듯이 하나의 개념에 머무르지 않고 끊임 없이 발전하고 변화하기 때문이다. 지금 정의를 내려도 미래에는 다른 형태로 발전되어 다른 모습일 가능성이 높다.

필자가 처음 클라우드에 대한 개념에 대해 깊이 생각해 본 계기는 2010년 초반 나왔던 미래학자 니콜라스 카가 쓴 빅스위치(Big Switch)라는 책을 통해서였다. 그는 책을 통해서 IT 시스템의 주류를 차지하던 자체 운영 환경(기업의 전산센터에 있는 대규모 컴퓨터 환경)은 마치 초기 전기의 발전과 발전소라는 중앙 집중식 리소스 관리 형태로 변할 것이라고 예상했다. 책에서 들었던 예는 전기였다.

1800년대, 에디슨의 전구 발명으로 사회적으로 확산되던 시기에 전기는 두 가지 관점에서 접근이 되고 있었다. 첫 번째는 안전하고 편리한 직류 전기를 각 가정과 공장에서 생산하는 모델이다. 마치 오늘날 IT 시스템을 직접 운영하는 모델과 같은 개념이다. 반면 당시 경쟁자였던 니콜라 테슬라(전기차 회사 테슬라의 이름을 따온 혁신가)는 교류 전기가 비록 위험하고 다루기는 어렵지만 발전소에서 중앙 생산해서 도시 전체에 공급하기 유리한 점 때문에 '발전소+교류' 형태의 모델이 유리하다고 생각했다. 그리고 결과는 아는 것처럼 테슬라의 모델이 승리했다.

오늘날 우리는 특별한 경우를 제외한다면 전기를 자체 생산하기보다는 발전소에서 중앙 공급받는 방식을 사용하고 있고 모든 산업과 생활이 이러한 비즈니스 모델에 맞추어 발전해

왔다. 직류는 감전이 되어도 치명적인 위험성도 없고 많은 전자기기에 바로 사용할 수 있다는 장점이 있음에도, 우리는 교류 전기를 받아 어댑터를 통해서 직류로 바꿔 사용하고 있다. 그것은 분명히 교류 기반의 비즈니스 모델이 훨씬 장점이 있다는 것을 의미한다. 사실 클라우드 컴퓨팅은 이러한 발전소+교류에 기반한 중앙 공급식 모델과 유사하다고 할 수 있다.

자 그러면 클라우드 또는 더 정확한 표현으로 클라우드 컴퓨팅이란 도대체 무엇일까? 다음은 위키피디아에 등록된 클라우드 컴퓨팅에 대한 정의이다.

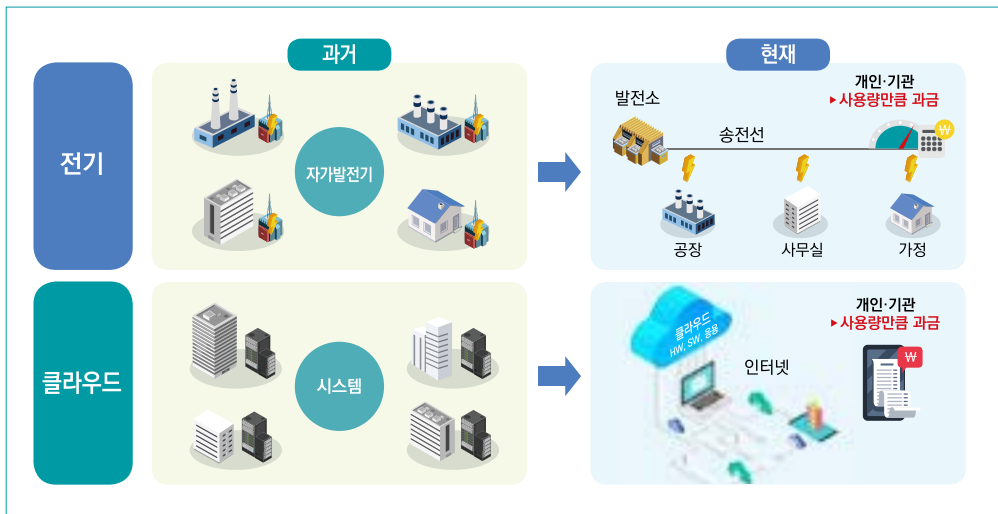
“클라우드 컴퓨팅(영어: cloud computing)은 사용자의 직접적인 활발한 관리 없이 특히, 데이터 스토리지(클라우드 스토리지)와 컴퓨팅 파워와 같은 컴퓨터 시스템 리소스를 필요 시 바로 제공(on-demand availability)하는 것을 말한다. 일반적으로는 인터넷 기반 컴퓨팅의 일종으로 정보를 자신의 컴퓨터가 아닌 클라우드에 연결된 다른 컴퓨터로 처리하는 기술을 의미한다.

... 중략 ...

클라우드 컴퓨팅은 전기망을 통한 전력망과 비슷한 일관성 및 규모의 경제를 달성하기 위해 자원의 공유에 의존한다.”

우선 여기서도 전기와 전력망과 비슷하고 비즈니스 모델이 유사하다는 얘기를 하고 있다. 그리고 여기서 사용하는 자원은 전기가 아닌 컴퓨팅 즉, 우리가 문서 작성이나 인터넷 쇼핑, 게임 플레이 등에 사용하는 자원을 사용한다고 말하고 있다. 클라우드가 동작하는 모델은 발전소에서 전기를 중앙 공급하듯이 중앙에서 자원을 관리하는 것이고, 다만 거기에 사용하는 자원이 컴퓨팅이라고 하는 IT 시스템이라고 생각하면 된다.

〈그림 11-1〉 전기와 클라우드 컴퓨팅 비교



출처: 과학기술정보통신부

다만, 필자가 생각하는 전기와의 차이는 전기는 단순한(물론 학문적으로는 복잡하겠지만, 일반인들이 느끼는 부분에서는) 반면, IT 시스템이나 컴퓨팅 자원은 전기보다 복잡하고 그것을 단순하게 중앙 공급이 가능할 것인가 하는 고민이 필요하다고 본다. 과연 전기처럼 가져다 쓰기만 하면 될까? 전기에서는 110V나 220V나 정도만 알면 되지만, IT 시스템에서는 컴퓨터, 네트워크, 데이터베이스 등 고려해야 할 부분들이 많다. 즉, 전기와 유사한 부분이 많이 있지만 우리가 클라우드에 대해 올바르게 알아야 할 상세한 부분도 있다는 의미이다.

이 책에서는 이러한 클라우드의 다양한 특성에 대해 설명을 할 것이다. 우선 클라우드는 발전소와 전기처럼 클라우드 서비스를 통해서 'IT 시스템이라고 하는 자원을 중앙 공급하는 모델이다'라고 이해하고 정리한다.

## 1-2 클라우드의 등장 과정과 배경

앞에서 클라우드 컴퓨팅은 전기와 비슷하다고 말을 했다. 하지만 두 가지는 다른 길을 걸어 왔다. IT 시스템은 전기와 달리 중앙 집중식으로 공급하는 모델과 분산형 모델을 번갈아가면서 사용하며 발전했다. 사실 1970년대까지는 PC라고 하는 개인용 컴퓨터 대신

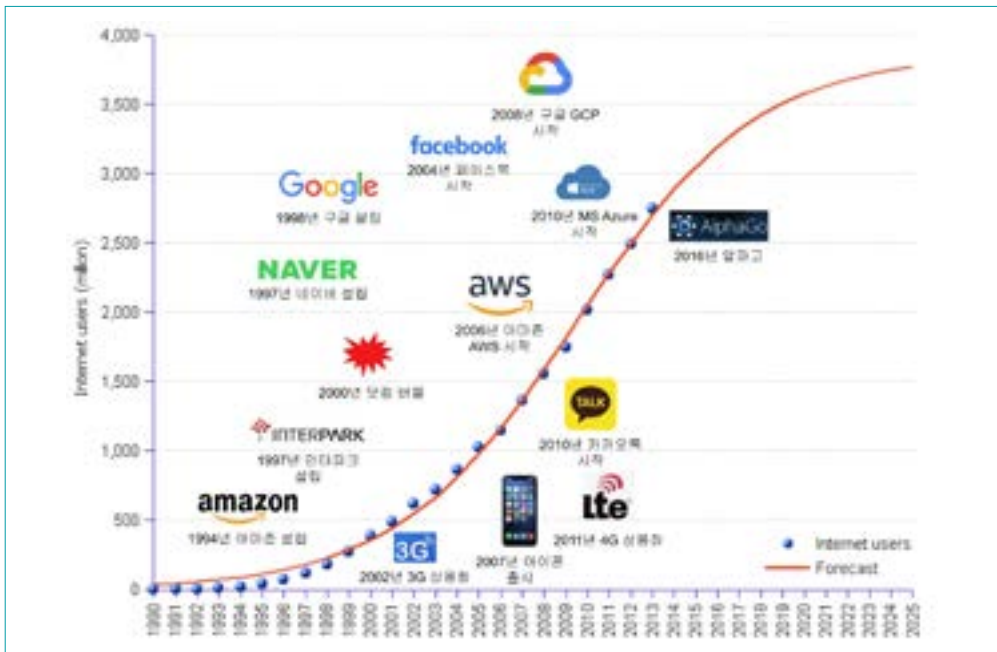
메인프레임이라고 하는 중앙 집중식 자원을 공유하는 방식을 더 많이 사용했다. 최초에 컴퓨터가 등장했을 때는 각 연구소에서 각자의 컴퓨터를 사용했지만, 산업이 발전하면서 필요한 컴퓨팅 파워(전기 용량이라고 생각하자)가 커지면서 분산형보다는 메인프레임 같은 대형 컴퓨터를 서로 같이 사용하는 방식이 훨씬 유리했다.

그러다가 PC 즉, 개인용컴퓨터가 급격히 발전하면서 각자 또는 각 단체가 자신들의 컴퓨터를 통해서도 충분한 용량을 사용할 수 있었다. 이러한 흐름은 1980년대부터 2000년대까지 이어졌다. 이 시기에 오늘날 우리의 생활을 혁신적으로 바꾸는 큰 사건이 있었다. 바로 인터넷의 등장이다. 지금은 너무나 자연스럽지만, 사실 인터넷은 우리의 생활을 인터넷 이전과 이후로 나눌 정도로 엄청난 변화였다.

[에피소드] 인터넷이 없던 시절 어떤 정보를 얻으려면 책이 주요한 매체였다. 책은 주로 서울의 대형서점에 가야 쉽게 구할 수 있었고, 수도권 도시에 살던 필자는 그러한 책을 사러 한 달에 한 번씩 종로에 있던 대형서점들을 자주 가야만 했다. 문제는 책을 사서 가지고 왔는데 정보가 불충분한 경우이다. 결국 다시 서점에 가서 추가로 서적을 구해와야 하고, 이는 다시 말하면 원하는 정보를 얻는 데 한 달 이상이 걸릴 수도 있다는 것이다. 오늘날 우리는 인터넷 서점에서 책을 사서 이러한 문제를 해결하기도 하지만, 사실은 아예 책이 아니라 블로그나 유튜브 등을 통해서도 충분한 정보를 빠르게 얻을 수 있다.

인터넷을 통해서 전 세계가 연결되기 시작하고 새로운 사업 모델도 무수히 등장했다. 인터넷 서점(아마존)에서 배달 서비스까지 우리의 삶 자체가 송두리째 바뀌었다. 문제는 그러한 변화를 뒷받침하기 위해서는 IT 시스템도 매우 복잡하고 큰 용량으로 바뀌었다는 것이다. 어느 날 인터넷 쇼핑몰에서 할인 행사를 하면 너무 많은 사람들이 모여 페이지가 뜨지 않는 경우도 있다. 유명한 국민 메신저 카카오톡이 안 되면 생활이 마비가 될지도 모른다.

〈그림 11-2〉 인터넷과 IT의 발전



출처: ITU(International Telecommunications Union)의 그래프에 직접 Event들을 추가하여 작성

컴퓨터 용량도 엄청나게 필요하지만 시스템도 복잡해져서 각 기업들이 해당 시스템 운영 인력을 보유한다는 것도 쉽지 않다는 의미이다. 인류는 반복적인 일이나 공통적인 업무가 있으면 그것을 하나로 통합해서 효율화하는 과정을 끊임 없이 해 왔다. 마찬가지로 IT 업계에서도 각자 이렇게 복잡한 시스템을 운영하는 것은 효율성도 낮을 뿐만 아니라 비용도 엄청나게 들어가는 일이라는 것을 깨달았다. 그리고 그것이 사회나 국가의 발전을 제한할 방해물이라는 것을 인식했다. 카카오톡이 마비되면 우리 생활이 불편해지는 것을 생각하자. 그래서 비록 전기보다는 수백 배, 수천 배 복잡할 수 있지만 그것을 중앙집중식으로 공급할 수만 있다면 엄청난 효율이 생긴다는 비전을 보고 등장하게 된 기술과 비즈니스가 바로 클라우드 컴퓨팅이다.

따라서 클라우드 컴퓨팅은 누군가의 아이디어에서 나온 발명품이라기보다는, IT를 기반으로 한 사회적인 발전 과정에 나온 현상이라 할 수 있다. 마치 산업혁명이 기계의 발전을 기반으로 사회를 변화시킨 것과 유사하다. 다만, 오늘날에는 산업혁명급의 현상이 자주

발생하는 것이고 그러다 보니 클라우드 컴퓨팅이라는 혁신적인 기술과 현상을 깊이 들여다볼 생각을 하지 못하는 것이다. 또 한 가지 IT를 전공하지 않은 일반인들이 클라우드를 체감하지 못하는 이유는 겉으로 잘 드러나지 않는 기반기술이기 때문이다. 초기 전기의 역사에서도 전기보다는 생활을 밝게 해주는 전구가 일반인들에게는 더 중요한 부분이었다.

### 1-3 플랫폼 비즈니스

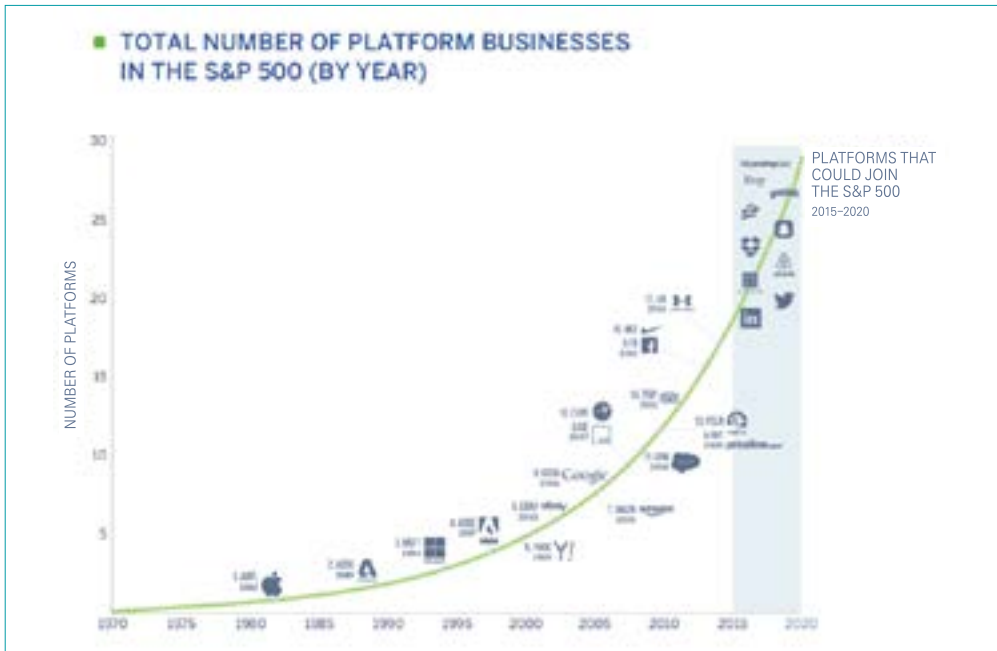
클라우드 컴퓨팅을 일반인의 입장에서도 엄청난 혁신이라는 것을 알게 해줄 만한 적절한 비유가 있다. 바로 유튜브와 같은 동영상 플랫폼 서비스이다. 클라우드는 IT 업계의 플랫폼 서비스 또는 플랫폼 비즈니스이다. 동영상 플랫폼이 우리에게 어떤 영향을 주고 있는지 한 번 생각해 보자.

유튜브와 같은 동영상 플랫폼이 없었던 시절 방송에 나갈 수 있는 사람들은 매우 제한적이었다. 아무리 좋은 콘텐츠를 가지고 있어도 방송에서 불러줄 정도의 인지도가 있지 않다면 대중들에게 매스미디어를 통해 알린다는 것은 쉽지 않았다. 그리고 그러한 방송은 전국적인 방송 네트워크를 가지고 있는 방송국에서만 가능했다. 동영상 플랫폼이 있는 요즘은 어떤가? TV에 나오지 않는 유명하지 않은 연예인들이나 일반인들 마저도 이러한 플랫폼을 통해서 방송을 진행할 수 있다. 콘텐츠를 다수에게 보내는 것이 어렵지 않다.<sup>20)</sup>

만약 이러한 현상들이 반복된다면, 과연 그 때도 사람들이 동영상 플랫폼의 현상을 방송에서 관심을 가지고 볼까? 굳이 동영상 플랫폼을 통해서도 계속 이어갈 수 있는 콘텐츠를 왜 방송으로 옮겨서 봐야 할까? 물론 극단적으로 방송이 사라질 것이라는 말을 하는 것이 아니다. 그만큼 플랫폼 비즈니스가 성장하고 있고 사회문화적으로 많은 영향을 주고 있다는 얘기를 과장해서 하는 것이다. 심지어 이러한 플랫폼들이 택시에서 음식배달에서 쇼핑에서 이미 업계를 움직이고 있다.

20) 글을 작성하는 지금 이 순간에도 브레이브걸스라는 데뷔 5년차 걸그룹이 크게 수익성이 나지 않는 군부대 공연만을 하다가 해체를 앞둔 시점에, 갑작스레 유튜브에서 유명세를 타면서 지금은 인기있는 걸그룹이 되었다. 되었다. 오히려 방송이 동영상 플랫폼의 현상(걸그룹의 유명세)을 뒤늦게 알리고 활용하는 수준이다.

〈그림 11-3〉 S&P 500에 포함된 플랫폼 비즈니스 기업들



출처: alleywatch.com

클라우드는 IT 플랫폼이다. 과거에는 아무리 좋은 비즈니스 모델이 있어도 대규모 IT 자원과 운영 인원이 없으면 사업을 하기가 어려웠다. 그러나 이제는 클라우드라는 플랫폼을 통해서 다양한 스타트업들이 본인들의 비즈니스 모델을 사업화하기 쉬워졌다. 물론 유튜브 콘텐츠 크리에이터들이 콘텐츠뿐만 아니라 유튜브라는 플랫폼을 잘 알면 본인들의 콘텐츠를 성공시키기 유리하다. 이것은 클라우드라는 플랫폼도 마찬가지이다. 새로운 비즈니스 특히 핀테크 기반의 새로운 비즈니스 모델을 고민하고 있다면, 왜 어느 정도는 클라우드를 알아야 하는지 이해할 수 있을 것이다.

## 2 기업들이 클라우드를 좋아하는 이유

### 2-1 클라우드 도입의 효과

앞에서 클라우드라는 것이 '중앙 집중식 IT 플랫폼 서비스'라는 것에 대해 알아보았다. 왜



여러 기업들이 클라우드에 대해서 관심을 가지고, 심지어 정부에서도 클라우드를 집중적으로 육성하려고 하는지 이해가 되었을 것이다. 기업의 관점에서 실질적으로 어떤 이득이 있을지 구체적으로 한 번 알아보자.

먼저 스타트업 입장에서는 마치 동영상 플랫폼처럼 비즈니스 모델만 있다면 사업을 시작하고 성공시키기 유리하다는 것을 알았다. 스타트업은 소수의 인원이 적은 자본을 가지고 작게 사업을 시작하는 것이다. 마치 콘텐츠만 가지고 방송을 시작하려는 크리에이터와 유사하다. 최근 트렌드를 보면, 단지 방송의 보조 수단이 아니라 방송이라는 생태계를 완전히 새롭게 정의하는 생태계 파괴자(disruptor를 번역한 파괴자라는 어감이 좋지 않지만, 혁신자라고 생각하면 좋을 듯)의 역할을 하고 있다. 마찬가지로 스타트업 입장에서 기존의 대기업 또는 전통적인 업계 강자가 지배하고 있는 업종에서 그들보다 유사하거나 오히려 더 좋은 IT 서비스를 제공할 수 있다. 즉, 비즈니스 모델만 우수하다면 그리고 물론 어느 정도의 자본도 뒷받침이 되어야 하겠지만 적어도 IT 시스템의 규모로 인한 경쟁 조건은 해결할 수 있다.

특히 스타트업들이 주로 하는 새로운 아이디어는 처음부터 성공 가능성을 예상하기 어려워 대규모 IT 투자에 위험이 따른다. 반면 클라우드를 사용하면 작게 시작(그만큼 작은 비용을 지불)해서 지속적으로 서비스를 개선해 나가며 사업을 키우는 것이 가능하다. 적어도 IT 시스템에서만은 기존의 강자와 동일한 조건에서 경쟁이 가능한 것이다. 따라서 스타트업들에게 클라우드 기반의 IT 서비스는 선택이 아니라 거의 필수라 할 수 있다.

전통적인 기업들, 주로 업계를 장악하고 있는 대기업 입장에서는 현재의 성공을 지속적으로 유지 또는 확대하고 싶은 목표가 있다. 그런데, 오늘날 대부분의 사업은 IT와의 결합을 생각하지 않을 수 없다. 택시도 IT 서비스로, 음식 주문도 IT 서비스를 통해서 하는 세상이다. 반면 선두주자의 입장에서 새로운 서비스 모델을 런칭하는 것은 쉬운 일이 아니다. 기존 서비스의 규모가 있다 보니 어느 정도 완성된 서비스가 나와야 하고, 또한 고객의 반응에 맞춰 빠르게 개선을 해야 한다. 즉, 소위 말하는 민첩성(agility)이 매우 중요하다. IT 기반 택시 서비스와 같은 새로운 트렌드가 등장했을 때, 빠르게 대응하면서 고객의 요구사항을

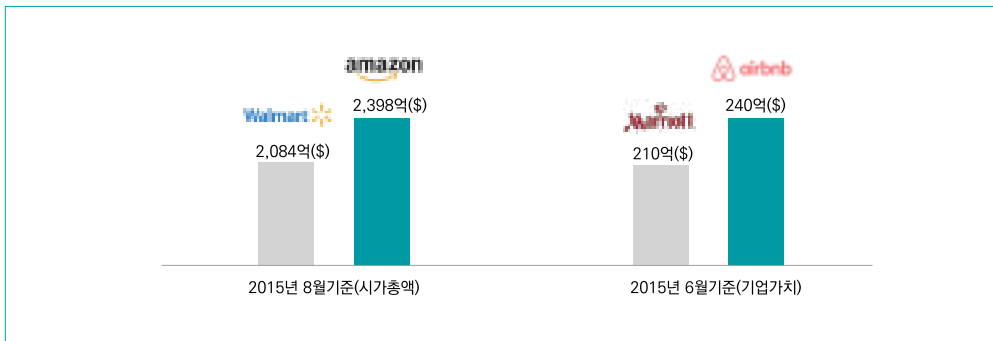
만족시켜야 할 필요가 있다. 그런 면에서 클라우드는 기존의 방식에 익숙한 전통적인 기업이 조직의 근간을 바꾸지 않고도 새로운 서비스를 작고 빠르게 런칭할 수 있는 장점을 제공할 수 있다.

후반부 클라우드 사례에서도 설명하겠지만 미디어 비즈니스를 시작하는 기업은 수많은 네트워크 인프라를 구축하지 않고도 클라우드 CDN서비스를 통해서 국내뿐만이 아니라 전 세계에 바로 원하는 콘텐츠를 서비스할 수 있다. 빅데이터나 인공지능 분야는 고도의 시스템 규칙이 필요하지만 클라우드에서 제공하는 시스템을 활용하면 설치 과정 없이 데이터만 있으면 바로 적용할 수 있다. 클라우드는 이렇게 다양한 서비스를 제공할 뿐만 아니라 무어의 법칙(24개월마다 반도체의 집적도가 2배로 증가한다는 인텔 공동창업자 고든 무어의 경험에 근거한 법칙)에 따라 기하급수적으로 증가하는 하드웨어에 대한 고정비용 투자를 할 필요 없이, 사용한 만큼 비용을 지불하는 pay-as-you-go 모델이 가능한 부분도 큰 장점이 아닐 수 없다.

## 2-2 디지털 트랜스포메이션

인터넷의 등장, IT 서비스의 발전, 플랫폼 기반 서비스 등을 통해 단지 IT나 비즈니스만이 아니라 사회 전체가 바뀌고 있다. 그에 맞추어 새로운 비즈니스 모델을 가진 혁신기업들이 속속 등장하면서, 업계의 전통적인 강자인 기업들에게는 민첩성이라는 숙제가 주어졌다. 디지털 트랜스포메이션은 기존의 물리적인 자원(기존 구매한 IT 시스템, 인력, 자본)을 클라우드 기반의 효율적인 서비스로 바꾸어, 기업들이 혁신적인 조직으로 변화하는 것을 말하는 용어이다. 즉, 디지털 기술을 통한 프로세스의 개선, 나아가 비즈니스의 혁신을 달성하는 것을 의미한다.

〈그림 11-4〉 전통기업과 혁신기업의 가치



출처: 디지털이니셔티브 그룹

개인적인 의견으로는 기업의 혁신이라는 목표에 필요한 클라우드(또는 디지털 기술)라는 수단을 다소 과장한 측면이 있다는 생각이 들지만, 반대로 보면 오늘날 여러 변화 속에서 혁신을 하기 위해 클라우드만큼 좋은 수단도 없다는 생각도 동시에 든다. 인터넷과 함께 또 다른 혁신을 가져온 스마트폰으로 인해 비즈니스 환경은 빠르게 변하고 있다. 혁신과 변화는 피해갈 수 없는 숙제이다. 심지어 넷플릭스 같은 혁신적인 기업들에게도 유사한 경쟁자들이 등장하고 유튜브 같은 동영상 플랫폼들과의 경쟁으로 또 다른 혁신이 필요한 상황이다. 이러한 상황에서 과거처럼 IT 기술 또는 시스템이 혁신의 발목을 잡아서 안 된다. 기업에게 당장 필요한 것은, 빠른 발전으로 만들어진 치열한 경쟁 상황에서 기업은 생존을 위해 혁신이다. 그러한 목적을 달성하기 위해 프로세스의 개선, 조직의 혁신 등이 필요하며, 그러한 부분을 클라우드 기반의 IT 서비스를 통한 조직의 변화 즉, 디지털 트랜스포메이션이 그 해답을 줄 수 있다는 공식인 셈이다.

### 3-1 금융 분야에서 클라우드

최근 금융 분야에서 클라우드에 대한 관심은 뜨겁다. 전통적으로 금융 분야는 IT 기술에 투자를 많이 하는 산업이다. 복잡한 금융 시스템을 안정적으로 유지하기 위해서는 IT에 대한 투자가 반드시 필요하기 때문이다. 그런데 최근 들어 이러한 복잡하고 안정적인 금융 IT 환경에 새로운 바람이 불었다. 그것은 핀테크 기술이 발전하면서 고객들이 편리한 유저 경험을 요구하기 때문이다. 대표적으로 카카오톡 메시지를 통한 결제, 송금 등을 처리해 보면 그동안 왜 우리는 어렵게 살아왔나 허탈한 감정이 들 정도로 정말 편리하다. 고객들은 이러한 환경을 접하게 되면서 최신 핀테크 및 IT 기술에 관심이 높아졌고, 전통적인 방식으로 시스템을 운영하던 금융 기업들에게도 편리한 방식의 금융 업무 처리를 요구하기 시작했다. 직접 금융사에 찾아가 처리하던 금융 업무를 인터넷뱅킹이나 HTS, 메신저 등을 통해 처리하는 비대면 방식에 대한 요구사항도 대폭 증가하였다. 물리적으로 움직이면서 발생하는 시간이 낭비스럽다고 생각하기 때문이다.

금융사들은 이러한 요구에 대응하기 위해 외부 업체들(핀테크 기업, 소셜 네트워크)과의 협업이 필요해졌다. 그러나 금융사는 안전한 금융거래를 위한 감독과 규제라는 태생적인 한계를 극복하는 것이 기존의 시스템을 가지고는 쉽지 않았다. 속도보다는 안전성이 중요하기 때문이다. 그러나 이러한 흐름은 시대적인 트렌드이고 안전성을 유지하면서도 변화의 속도를 올리는 것은 당면과제가 되었다. 클라우드는 이런 부분에서 개방성과 민첩성이라는 장점을 부여하고 있다. 특히 데이터 분석이나 AI/ML 같은 최신 기술처럼 현대 기업들의 경쟁력에 중요한 요소 기술들을 쉽게 확보할 수 있다. 외부 서비스와 연계를 할 때도 클라우드를 이용하면, 단순히 기술적인 장점만이 아니라 클라우드가 가진 넓은 생태계를 통해 사실상 표준 기술을 사용하여 개방성을 극대화할 수 있다. 클라우드를 활용하면 전통적인 금융 기업들이 최신 트렌드를 쫓는 핀테크 기업들과의 협업이 용이하다. 이는 결국 금융 기업들로 하여금 금융상품이라는 본연의 업무에 충실하면서도 경쟁력을 높일 수 있는 좋은 방법이 되었다.

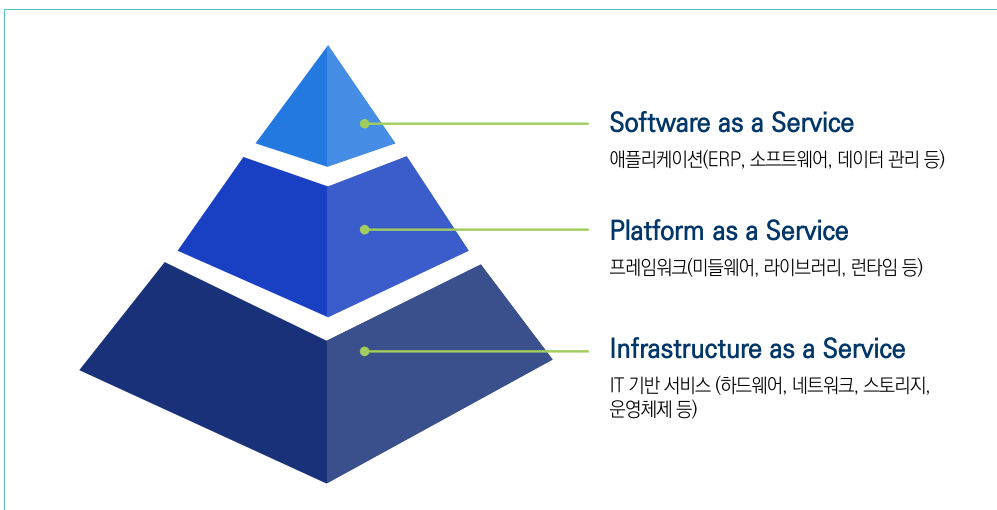
### 3-2 핀테크와 클라우드

핀테크 기업들은 기존 금융사의 업무 처리 방식에 새로운 기술을 접목하여 현대적인 서비스를 제공하고 있다. 오픈 API 기반의 서비스를 통해 금융을 일종의 플랫폼 비즈니스로 제공하기도 한다. 개인들의 취향에 맞는 서비스를 제공하기 위해 빅데이터 분석을 하고 인공지능 기술을 접목한다. 보안에 있어서도 외부 침입 차단과 같은 기본적인 부분은 물론 인공지능 기술을 활용한 패턴 감지와 같은 스마트한 기능들을 구현한다. 다양한 금융사들의 전통적인 금융 업무를 빠르고 쉽게 시장에 제공하기 위해서는 발전된 IT 기술을 접목하는 것이 매우 중요하다. 상대적으로 새롭고 모험적인 비즈니스 모델이다 보니, 성공 여부가 불투명한 서비스를 작게 시작하여 점진적으로 확장해 가는 리스크 회피 모델도 필요하다. 클라우드는 핀테크를 시작하는 스타트업들에게 최적의 환경을 제공한다. 종량제 모델로 작게 시작하여 원하는 수준까지 빠르게 확장할 수 있는 기능을 제공한다. 그럼에도 불구하고 최신 IT 기술들을 완성도 높게 구현하여 제공하기 때문에 작은 리스크로 도입할 수 있다.

## 1 서비스 모델별 분류 (IaaS, PaaS, SaaS)

클라우드란 매우 광대한 개념이고 앞으로도 계속 그 범위나 의미도 비즈니스의 발전에 따라 조금씩 바뀔 것이라고 언급했다. 마치 자동차가 처음 발명될 때는 마차를 대신하는 운송 수단 정도로 인식이 되었지만, 현재는 승용차, SUV, 트럭, 버스 등으로 세분화되면서 발전한 것과 유사하다. 만약 여러분들이 자동차를 구입한다면 용도에 맞게 선택할 것이며, 가족들의 운송 수단을 위해 트럭이나 버스를 구매하지는 않을 것이다. 마찬가지로 광대한 클라우드의 개념을 이해하고 비즈니스 목적에 맞는 것을 고르기 위해서는 그것을 분류해 보면 좋다. 먼저 클라우드는 서비스를 제공하는 모델에 따라 IaaS, PaaS 그리고 SaaS로 구분한다.

<그림 II-5> IaaS, PaaS, SaaS

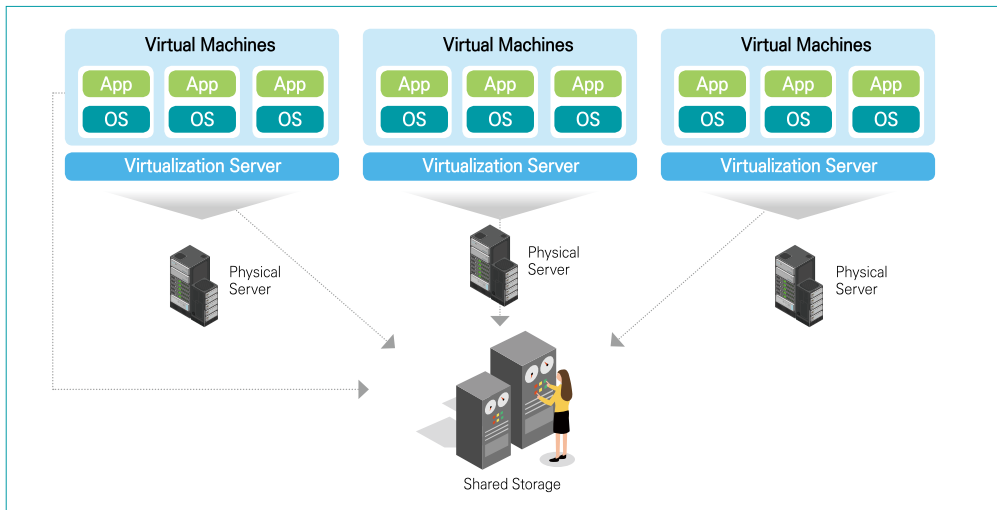


## 1-1 IaaS(Infrastructure as a Service)

건물을 짓는다고 생각해 보자. 사실 건물을 짓는 것은 IT 시스템을 만드는 것과 매우 유사하다. 건물을 짓기 위해서는 먼저 토지를 준비하고 기반을 다지는 것이 중요하다. 특히, 큰 건물을 짓기 위해서는 기초가 튼튼해야 한다. 클라우드에서 IaaS(이아스 또는 아이아스로 발음)는 건물의 기반과 동일한 의미인 Infrastructure를 의미한다. IT 시스템은 기본적으로 연산처리를 위한 컴퓨팅 자원(CPU)과 데이터를 저장하기 위한 스토리지(디스크), 그리고 여러 자원들을 연결하기 위한 네트워크가 필요하다. 이렇게 컴퓨팅, 스토리지, 네트워크 등의 기본적인 자원들을 그대로 제공하는 클라우드 서비스를 IaaS라고 한다.

보통 클라우드 공급자(아마존, 구글)들은 이러한 기본적인 자원들을 물리적으로 대량 보유하고 있다. 그리고 그것을 사용자에게 나눠준다. 다만, 물리적으로 보유한 자원을 그대로 사용자에게 분배하려면 효율성이 떨어질 수 있어 보통 가상화(virtualization)라는 기술을 사용한다. 가상화는 물리적인 단위로 쪼개져 있는 자원들을 하나의 거대한 풀(pool)로 만들어 논리적으로 분배하는 역할을 한다. 예를 들어 16코어 CPU 장비를 10대 보유하고 있다면, 가상화를 통해 160코어 풀(pool)로 만들고 거기에서 2코어, 4코어, 8코어 등 원하는 크기로 나눠 제공한다. 그러면 2코어 사용자를 위해 16코어 물리 장비 전체를 할당하는 낭비를 방지할 수 있다. 사실 클라우드 사용자는 특별한 경우가 아니라면 이렇게 가상화를 통해 제공된다는 부분을 이해하지 않아도 된다. 4코어를 신청했다면, 4코어 장비(실제로는 가상 장비)를 받아서 사용한다고 간주해도 무방하다. 클라우드 업체들은 효율성을 위해 CPU만이 아니라 디스크, 네트워크 등도 가상화해서 관리하고 있다.

〈그림 11-6〉 가상화 기술



클라우드 사용자가 IaaS를 통해서 컴퓨팅 자원을 할당받았다면, 그 위에 운영체제를 설치(보통은 클라우드 업체에서 설치해서 제공)하고 필요한 소프트웨어를 설치 또는 개발해서 서비스를 구축하면 된다. 건물에 비유한다면 기초작업이 된 토지를 받아, 그 위에 건물을 짓는 것과 유사하다고 할 수 있다. IaaS는 기본적인 자원만 할당받아 사용하므로 그 위에 어떤 것이든 원하는 형태로 구현할 수 있다는 장점이 있다. 반면, IaaS 위에 필요한 서비스를 구축하는 능력이 부족하거나 IaaS 자원을 관리할 인원이 부족하다면 어려움을 겪을 수 있다.

## 1-2 PaaS(Platform as a Service)

IaaS는 마치 기초가 다져진 토지를 받아 그 위에 건물을 짓는 것과 유사하다고 했다. 그러나, 만약 여러분들이 건물 안에 예쁜 상가를 내고 싶은 것이라면 건물부터 짓는 일은 하지 않는다. 대신 이미 만들어진 신축 건물에서 적당한 위치와 크기를 가진 한 칸을 빌리거나 구매하여 인테리어를 할 것이다. PaaS는 건물이라는 플랫폼을 제공하는 서비스라고 생각하면 된다. PaaS를 사용하면 이미 서비스 구현에 필요한 어느 정도의 소프트웨어가 구축되어 있고, 클라우드 사용자는 그 위에 원하는 서비스를 개발하여 사용하면 된다. 이러한 PaaS 서비스는



각 클라우드마다 조금씩 다른 모습을 하고 있는데, 대체로 데이터 분석 서비스, AI 서비스, 애플리케이션 서비스, 미디어 서비스 등을 공통적으로 제공한다.

여러분들이 넷플릭스와 같은 미디어 서비스를 한다고 가정해 보자. 그러면 많은 영상 콘텐츠들을 관리하고 스트리밍으로 서비스하는 기능이 필요할 것이다. IaaS를 사용한다면 컴퓨팅 자원(가상화하여 제공하는 컴퓨터라 VM 또는 Virtual Machine이라고도 함), 디스크, 네트워크를 할당받아 그 위에 콘텐츠 관리 및 스트리밍 서비스 소프트웨어를 개발해야 한다. 하지만, 여러분들은 고객들에게 영상 서비스를 하고 싶을 뿐이지 미디어 관리나 스트리밍 서비스 개발에 관심이 없을 수도 있다. 사실 대부분의 스타트업들은 기반 IT 기술보다는 고객 대상 서비스에 관심이 있다. 이런 경우 클라우드 업체가 제공하는 미디어 PaaS 서비스를 사용하게 되면, 영상 관리나 스트리밍 서비스가 자동으로 제공이 된다. 그러면 여러분들은 이러한 영상 관리에 필요한 화면(User Interface)만 원하는 형태로 개발해서 연결하면 된다.

이외에도 빅데이터 분석을 원한다면 데이터 분석 서비스를 사용할 수도 있다. IaaS 위에 데이터 분석 서비스를 위해 필요한 소프트웨어를 설치 또는 구현하는 것은 매우 어려운 일이다. 특히 소규모 인원으로 시작하는 스타트업에서 이런 일은 거의 불가능에 가깝다. 이런 경우는 PaaS 서비스를 사용하는 것이 훨씬 유리하다. 요즘 클라우드에서는 PaaS 기반의 AI 서비스도 제공해서 영상인식(vision)이나 번역(translation), 음성인식(speech recognition) 등의 기능을 쉽게 구현할 수도 있다.

### 1-3 SaaS(Software as a Service)

마지막으로 SaaS는 아예 클라우드가 제공하는 서비스를 통해 비즈니스를 구현하는 경우이다. 건물에 비유하자면 이미 인테리어까지 끝난 상가를 제공하는 것이라고 보면 된다. 앞에서 예를 들었던 넷플릭스와 같은 미디어 서비스라면, 마치 유튜브 플랫폼에 본인의 채널을 만들고 영상을 올려 제공하는 것이라고 생각하면 된다.

여기서 의문이 들 것이다. 여러분들이 원하는 상가가 원하는 형태대로 인테리어가 된 것이

아니라면? 의류 매장을 원했는데 식당으로 인테리어가 되어 있다면? 유튜브로 동영상만 제공하는 것이 아니라 페이스북과 같은 소셜네트워크 서비스를 원한 것이라면? 넷플릭스 같은 드라마, 영화 서비스를 제공하고 싶은 것이었다면? 맞다. SaaS는 이미 고객이나 사용자가 바로 사용할 수 있는 형태의 서비스이기 때문에 여러분들이 원하는 목적과 잘 맞아야 한다. IaaS나 PaaS에 비해 이미 완성된 서비스를 제공하는 것인 만큼 비즈니스에 맞게 변경(customization)하는 것에 제약이 따른다. 반면, 식당으로 인테리어되어 있지만 카페를 하는 경우 정도라면, 유튜브의 영상 페이지를 가져다 홈페이지에서 보여주기만 해도 된다면 어느 정도 가능하다.

IaaS, PaaS, SaaS 서비스는 각각의 활용 범위와 완성도 면에서 다른 것이다. IaaS는 어떤 것도 그 위에서 할 수 있지만, 완성도는 낮기 때문에 거의 모든 것을 추가로 구현해야 한다. 반면 SaaS는 완성된 서비스이므로 추가로 구현할 것이 거의 없지만, 대신 이미 만들어진 기능을 변경하는 것이 매우 어렵다. 여러분들이 많이 사용하는 구글의 gmail 서비스도 일종의 SaaS라고 보면 된다. 요즘은 이메일 서비스를 직접 구현해서 사용하는 경우는 거의 없을 것이다.

어떤 경우에 IaaS를 사용하고 어떤 경우에 PaaS, SaaS를 사용해야 할까? 우선 SaaS는 완성도가 높으므로 여러분들이 구현하려는 비즈니스의 목적에 맞는 것이 있는지 우선적으로 SaaS 서비스들을 검토해 보고, 어느 정도 비슷하게라도 요건을 충족하는 것이 없다면 빨리 포기하고 IaaS나 PaaS 기반으로 구축해야 한다고 보면 된다. IaaS는 조직 내에 IT를 능숙하게 다루는 전문가가 있어야 한다. 클라우드 초기에는 PaaS나 SaaS로 가능한 선택지가 부족해서 IaaS를 많이 사용해야 했지만, 지금은 그렇지 않다. 특히 PaaS의 경우 무조건 어느 정도의 자원을 확보해야 하는 IaaS와 달리, 필요한 만큼의 자원만 사용할 수 있도록 설계되어 있기 때문에 비용 면에서도 유리하다. 따라서 대부분의 경우 PaaS가 유리하다고 볼 수 있다. 다만, PaaS는 클라우드 공급자마다 다소 차이가 있기 때문에, 한 번 선택을 하면 다른 공급자로 옮겨가는 것이 IaaS에 비해 어렵다는 점은 주의해야 한다.

〈표 II-1〉 IaaS, PaaS, SaaS

구분	IaaS	PaaS	SaaS
활용 난이도	높음	중간	낮음
완성도	낮음	중간	높음 <sup>주1)</sup>
활용 범위	무제한	거의 무제한 <sup>주2)</sup>	특정 서비스
구축/운영 비용	높음	약간 높음 <sup>주3)</sup>	낮음
사용료	높음	중간	중간 <sup>주4)</sup>
공급지 변경	쉬움	약간 어려움	거의 불가능

주1) SaaS의 가장 큰 장점은 완성도가 높아 구축에 들어가는 시간이 거의 필요 없다는 점이다.

주2) PaaS의 범위가 넓은 편이라 필요한 경우 IaaS처럼 원하는 서비스를 입맛에 맞게 구현 가능하다.

주3) 완성형 SaaS 서비스에 비해, PaaS는 구축이나 운영에 IT 인력(비용)이 어느 정도 필요하다.

주4) SaaS의 초기 소규모 사용시 비용은 IaaS, PaaS에 비해 절대적으로 낮지만 장기적으로 규모가 커지면 상승할 수 있다. 특히, SaaS는 불필요한 기능까지도 사용해야 하는 경우가 있다.

## 2 서비스 제공자별 분류(Public, Private, Hybrid, Community)

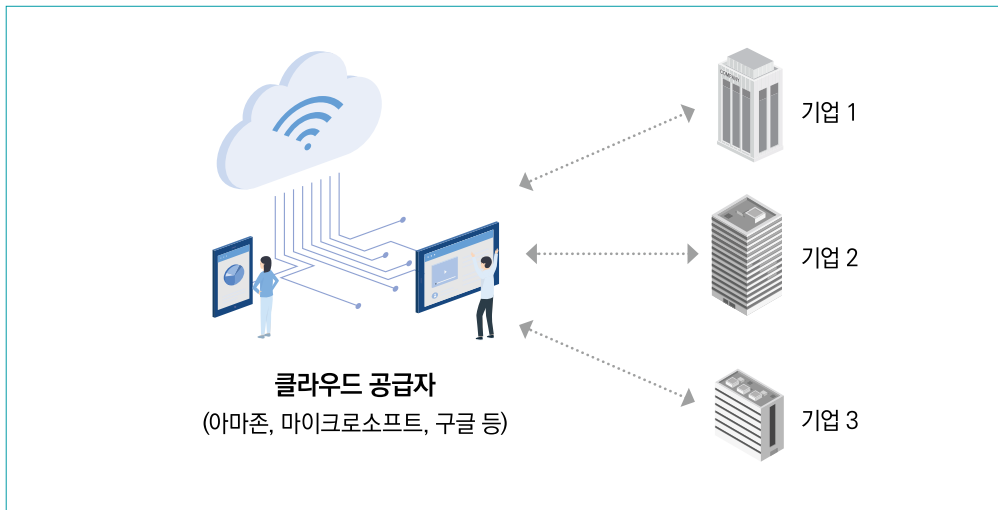
앞에서 클라우드 서비스를 제공하는 방식에 따라 IaaS, PaaS, SaaS로 나누어 보았다. 이번에는 서비스 제공자 또는 구축 방식(deployment model)에 Public, Private, Hybrid 그리고 Community 클라우드로 나누어 알아본다.

### 2-1 Public 클라우드

Public 클라우드는 아마존, 구글 등과 같은 클라우드 공급자가 대규모로 서비스를 구축하여 클라우드를 사용하려는 기업들에게 제공하는 방식이다. Public 클라우드는 마치 발전소에서 전기를 생산하여 공장이나 가정에 공급하는 것과 마찬가지로, 여러 기업들을 한 곳에서 서비스하므로 규모의 경제를 달성할 수 있어 비용 면에서 유리하다. 특히 클라우드 사업자는 대규모 IT 서비스 플랫폼을 지속적으로 운영하므로 많은 노하우와 인력을

보유할 수 있다. 따라서 기업 입장에서 Public 클라우드를 사용하게 되면 저렴한 비용만이 아니라 최상의 서비스의 품질이나 최신 기술을 제공받을 수 있는 장점이 있다. 일반적으로 클라우드라고 하면 Public 클라우드를 가리키는 경우가 많다.

〈그림 11-7〉 Public 클라우드 서비스 모델



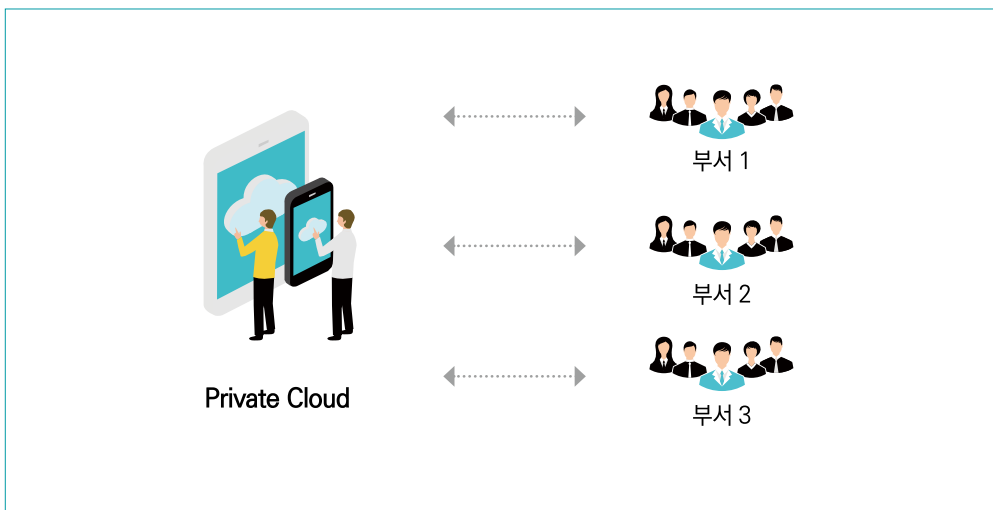
이러한 장점에도 불구하고 Public 클라우드는 데이터를 기업 외부에서 처리해야 하기 때문에 해당 업종이 민감한 데이터에 대한 규제를 받는 분야(금융, 의료)인 경우에는 조심을 해야 한다. 클라우드 공급자의 IT 자원은 여러 국가에 위치한 데이터센터에 보관이 되기 때문에, 어떤 경우는 해외 데이터센터에서 데이터가 처리되는 경우도 있다. 그러므로 이러한 업종에서 클라우드를 사용할 때는 법규상 문제가 없는지와 데이터의 저장 또는 처리 위치가 국내인지 해외인지 확인해야 한다.

## 2-2 Private 클라우드

이런 규제를 받는 분야나 자체적으로 대규모 IT 자원을 사용하는 기업의 경우에는 Public 클라우드 대신 기업 내부에서 클라우드 서비스를 하는 Private 클라우드를 구축하기도 한다.

Private 클라우드는 기업 내부의 여러 부서나 프로젝트에서 필요로 하는 IT 자원을 Public 클라우드처럼 한 군데 모으되 대신 기업 내부에 위치한 데이터 센터에서 처리하는 방식이다. 마치 전기 분야에서 발전소 대신 공장 내부에 발전기를 설치하고 전력을 생산하는 것과 유사한 방식이다. Private 클라우드는 아무리 기업의 규모가 크다고 하더라도, Public 클라우드 공급자만큼의 규모의 경제를 달성하기가 어려워 비용적으로 불리하다. 또한 기업 자체적으로 클라우드를 운영해야 하다 보니 대규모 인력을 보유해야 하고 특히 빅데이터나 AI/ML과 같은 최신 기술을 적용하기 위해서는 많은 투자가 필요하다. 그럼에도 불구하고 기업의 중요 데이터를 외부로 내보낼 필요가 없기 때문에 일부 기업들이 사용하고 있는 방식이다.

〈그림 II-8〉 Private 클라우드 서비스 모델

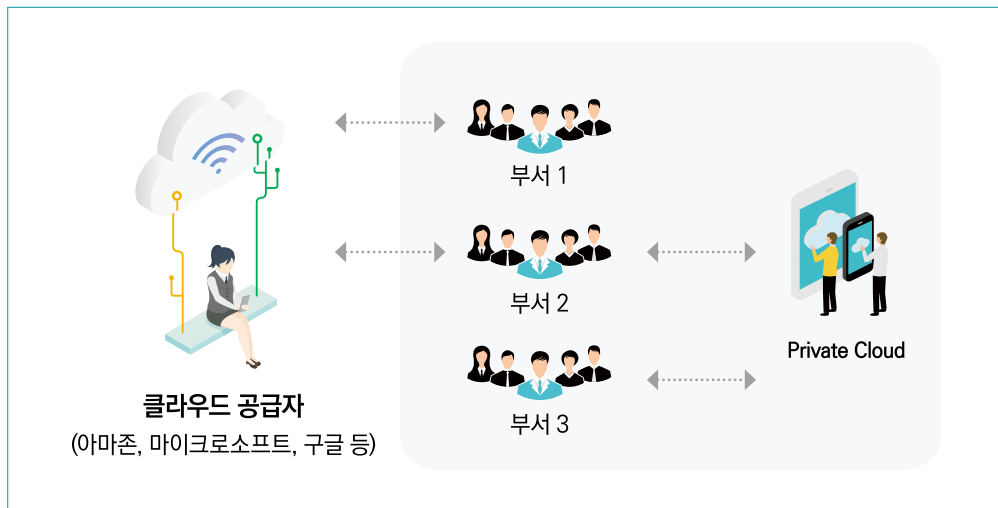


Private 클라우드의 경우에는 Public 클라우드에 비해 규모가 작기 때문에 IaaS, PaaS, SaaS를 모두 제공하기보다는 IaaS와 부분적으로 PaaS 서비스를 제공하는 형태가 가장 일반적이다. Private 클라우드는 기존에 사용하던 물리적인 하드웨어 기반의 IT 시스템에 가상화된 자원 관리를 추가하여 자원 관리를 효율화한다.

## 2-3 Hybrid 클라우드

Public 클라우드와 Private 클라우드는 각각 장단점이 명확하다. 반대로 기업의 입장에서 모든 IT 서비스들이 같은 성격을 가지고 있지는 않다. 즉, 일부 중요 업무는 민감 데이터로 인해 Private 클라우드가 필요하지만 그렇지 않은 업무는 Public 클라우드를 통해서 비용절감과 최신 기술을 빠르게 적용할 수 있는 장점을 얻을 수 있다. 따라서 기업의 입장에서는 Public이나 Private 중에서 하나를 사용하지 않고 업무별로 두 가지 중에서 하나 또는 두 가지를 모두 사용하는 방식을 Hybrid 클라우드라고 한다. 새로 비즈니스를 시작하는 스타트업의 경우에는 Public 클라우드에 모든 IT 서비스를 운영하는 경우가 많겠지만, 클라우드가 활성화되기 이전부터 비즈니스를 해온 기업들 입장에서는 모든 서비스를 Public 클라우드로 한 번에 옮기는 것은 쉽지 않다. 따라서 사실상 Public 클라우드를 사용하는 대다수의 기업은 Hybrid 클라우드를 사용한다고 볼 수 있다.

〈그림 11-9〉 Hybrid 클라우드 서비스 모델



Hybrid 클라우드는 Public과 Private의 장점을 골고루 활용할 수 있다. 특히 Public 클라우드의 비용 효율성 및 최신 기술을 활용하면서도 Private 클라우드를 이용하여 데이터를 기업 내부에 보관할 수도 있다. 다만, 클라우드 서비스를 사용하는 IT 팀 입장에서는 Public과 Private 클라우드 두 가지 환경을 모두 관리하면서, 둘 사이에 효율적인 연계(데이터

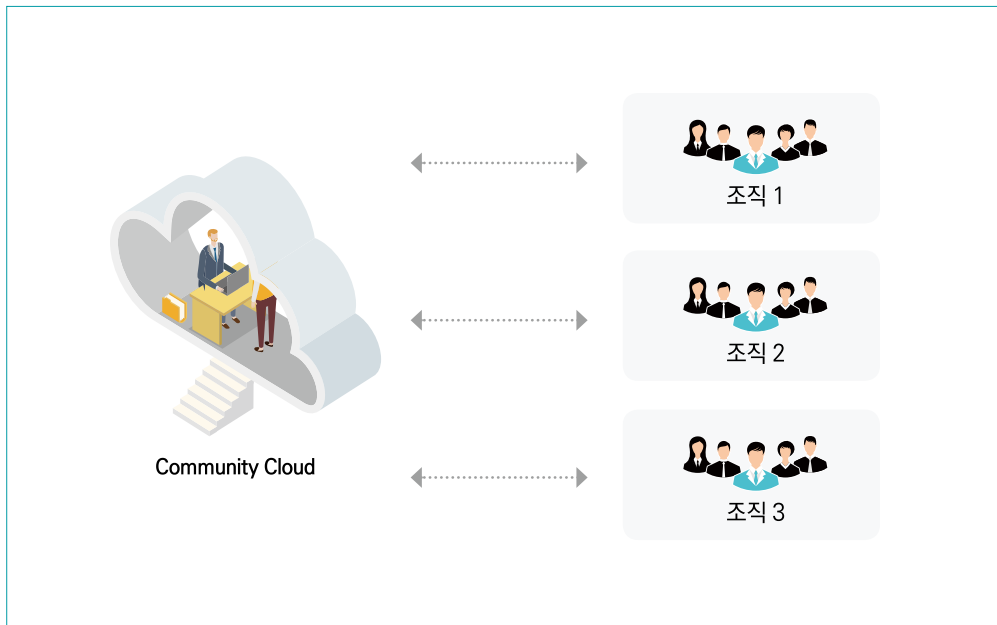
이동, 네트워크 연결)를 해야 하는 숙제가 있다. 예를 들어 Private에서 운영 중이던 서비스를 Public으로 이전할 때 환경적인 차이로 인한 호환성 문제를 극복할 수 있도록 시스템을 구현해야 하는 어려움이 생긴다. 그럼에도 불구하고 Hybrid 기술은 클라우드가 발전하면 발전할수록 기업의 입장에서는 효율성을 위해 반드시 고려해야 할 대상이다.

Hybrid 클라우드를 말할 때 또 한 가지 빼놓을 수 없는 경우는 바로 복수의 클라우드들을 비즈니스 목표와 서비스별 특성에 맞게 도입하는 멀티 클라우드이다. 기업의 입장에서 자사의 데이터센터를 활용한 Private 클라우드가 서비스를 운영할 수 있는 하나의 환경이라면, 복수의 Public 클라우드는 제2, 제3의 환경을 제공하는 셈이 된다. 멀티 클라우드를 도입하는 주요 이유는, 클라우드 공급자별로 강점을 가지는 서비스 영역(예를 들어 IaaS에 강한 A 클라우드, 데이터 분석에 강한 B 클라우드)이 다르기 때문에, 기업 입장에서는 취사선택을 통해 가장 효율이 높은 Best-Of-Breed(여러 가지 종류의 소프트웨어들을 조합하여 최상의 환경을 만듦) 환경을 만들 수 있다는 점이다. 제4절에 나오는 대부분의 사례들이 복수의 클라우드를 서비스별 특성에 맞게 도입한 멀티 클라우드 사례이다.

## 2-4 Community 클라우드

Community 클라우드는 Private 클라우드의 한 형태라고 보는 것이 좋을 듯 하다. 기업 내부에 클라우드를 구축하는 모델을 Private 클라우드라고 했는데, 만약 여러 기업(조직)이 동일한 법이나 규정을 공유하는 경우라면 굳이 각자 클라우드를 구축하지 않고 하나를 구축해서 여러 조직이 공유하는 모델이다. 주로 대기업에서 여러 계열사를 거느리고 있는 경우에 IT 계열사에서 클라우드를 구축해서 제공하는 경우가 이에 해당된다. 또는 정부의 관리 감독하에 비즈니스가 운영되는 금융 업종사들을 하나로 묶어 일종의 '금융 클라우드'를 제공하는 것도 Community 클라우드라고 할 수 있다. Private 클라우드에 비해 다수의 기업들을 대상으로 하므로 어느 정도 규모의 경제를 달성할 수 있고, 유사한 법이나 규정을 공유하다 보니 그에 특화된 클라우드 서비스를 맞춤형으로 제공할 수 있다는 장점이 있다.

〈그림 II-10〉 Community 클라우드 서비스 모델



### 3 핀테크에 맞는 클라우드 서비스는?

IaaS, PaaS, SaaS 그리고 Public, Private, Hybrid, Community 클라우드 중에서 어떤 것이 핀테크에 적합할까? 사실은 어떤 특정 모델이 핀테크에 적합하다고 말하기는 어렵다. 핀테크는 금융과 IT의 결합이라는 점에서 공통 분모를 가지고 있지만, 다양한 아이디어를 기반으로 한 비즈니스이기 때문이다. 따라서 핀테크에 적합한 클라우드 모델을 선택하는 과정은 핀테크로 하고자 하는 비즈니스를 이해하는 과정이 될 것이다. 이 책에서는 핀테크 자체에 대한 자세한 내용은 다루지 않을 것이므로 여기서는 예상되는 일반적인 핀테크 비즈니스 모델을 간략히 살펴보고 그에 맞는 클라우드 모델이 무엇인지 짚어보려고 한다.

당장 떠오르는 핀테크 비즈니스 모델로는 다음과 같은 것들이 있다. 혹시 빠진 게 있다면 여기서 설명하는 사례들을 바탕으로 제시하는 클라우드 모델을 응용해서 적용해 보기 바란다.



- 대부 중개, 클라우드 펀딩 및 P2P 대출 - 금융기관 연동, 빅데이터 분석, 인공지능
- 지급결제 및 전자결제 - 금융기관 연동, 보안 인증, IoT
- 개인 자산관리 - 오픈 API, 빅데이터 분석
- 자금이체 및 송금 - 금융기관 연동, 사용자 인증
- 디지털 화폐 - 블록체인
- 금융 업무 프로세스 개선 - 빅데이터 분석, HPC
- 인슈어 테크 - IoT

### 3-1 대부 중개, 클라우드 펀딩 및 P2P 대출

전통적인 대부업은 금융 고객의 신용도를 기반으로 자금을 대여하고 그에 대한 이자를 수익으로 가져가는 업종이다. 그런데 자금을 빌려주는 주체가 인허가를 받은 금융사들을 통해서만 이루어지다 보니 1, 2금융권에서 대출을 받기 어려운 사람들에게는 필요한 자금을 융통하는 것이 어렵거나 높은 이자를 물어야 한다. 핀테크에서는 데이터 기반으로 신용도를 조금 더 정밀하게 측정하여 대출이 어려운 고객들 중 상환 능력이 있는 사람들에게 대출을 해주는 것이라고 보면 된다. 그리고 인허가 기관이 아닌 개인들끼리 대출이 가능하도록 중개하기도 한다. 문제는 기존의 신용도 평가를 어떻게 개선할까 하는 것이다. 기존의 신용도 평가 방식은 대출자의 자산, 직장, 급여 등을 기준으로 판단한다. 반면, 데이터 분석 기반의 신용도 평가는 여기에 대출자의 금융 거래, 구매 이력, 운영 중인 사업체의 매출 트렌드, 사업체의 현금 흐름, 소셜네트워크 활동 내역 등까지 추가하여 판단한다. 특히, 기존의 방식이 이미 결정된 과거의 데이터(실제 자산 가치가 지표상 반영될 때까지 지연)를 기준으로 삼는 반면, 실시간 발생하는 데이터를 활용하여 적시성을 높일 수 있다.

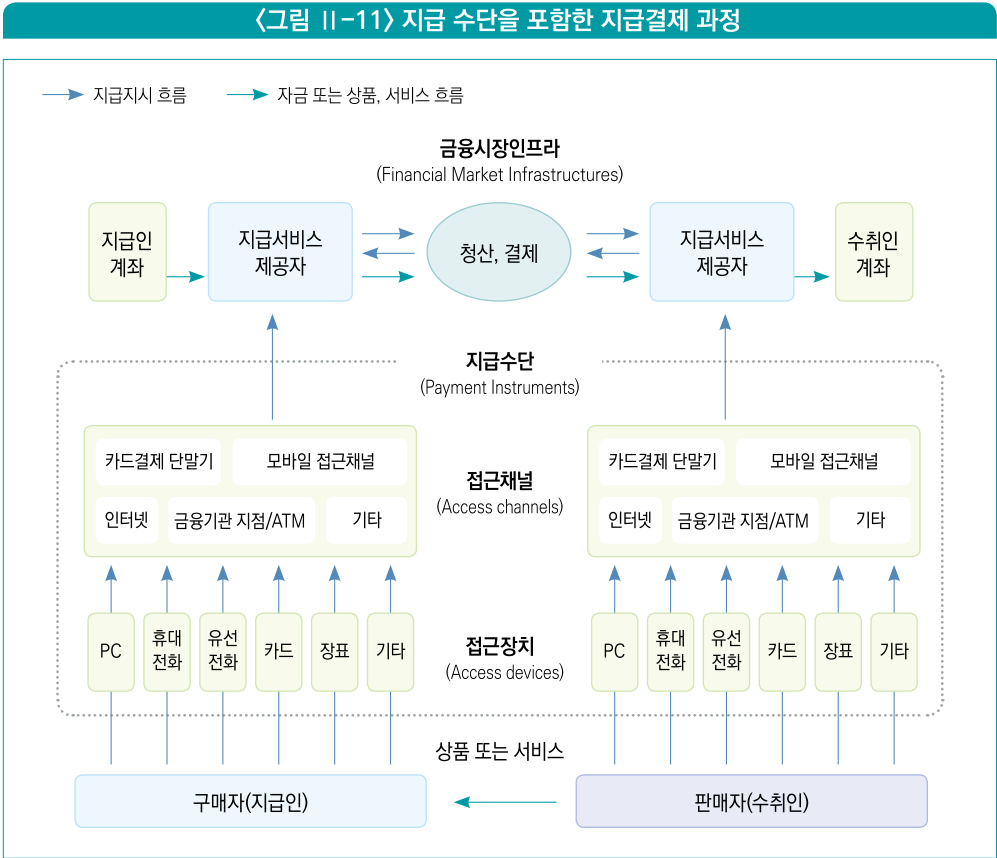
여기에 필요한 기술은 대략 여러 가지 데이터를 제공할 수 있는 기관이나 서비스와 연동, 수집된 데이터에 대한 분석, 인공지능 기반 예측 등이다.

### 3-2 지급결제

지급결제는 원하는 서비스나 상품을 구매후 구매자(지급인)가 판매자(수취인)에게 재화의 구매 대가를 지불하는 과정에 지급인과 수취인 사이에 거래를 승인하고 비용을 지불하는 중계 서비스를 제공하는 분야이다. 먼저 구매자가 구매를 위해 결제에 필요한 스마트폰, 카드 등의 장치(device)를 이용하여 지불 의사를 밝힌다. 그러면 판매자는 해당 서비스나 상품을 제공하고 거래 내역을 결제 서비스 제공자에게 전달하여 수신 의사를 밝힌다. 그러면 결제 서비스 제공자가 판매자의 계좌에서 자금을 인출하여 판매자의 계좌로 그것을 전달하여 결제 과정을 마친다. 이는 전통적인 금융 거래에서와 핀테크가 다를 바 없다. 다만, 핀테크를 통해 주로 개선하는 부분은 결제에 필요한 장치를 다양화하여 구매자나 판매자가 거래를 보다 편하게 하는 것이다. 예를 들어 최근에 많이 사용되고 있는 카카오페이와 삼성페이의 경우 각각 카카오톡 메신저와 삼성 스마트폰을 이용하여 거래를 가능하게 하여 카드를 가지고 다니는 불편함을 해소하였다.

삼성페이 같은 경우에는 스마트폰 디바이스에서 카드 단말기와 통신하는 IoT 기술이 먼저 필요하다. 카카오페이는 메신저상의 거래에서는 기기가 불필요하므로 온라인에서 금융기관과 통신하고 구매자를 인증하는 기술이 필요하다. 오프라인 거래에서는 삼성페이처럼 IoT 기술을 활용하지 않고 카메라를 통해 QR 코드를 인식하는 기술을 활용한다. 아무래도 카카오페이용 QR 코드 단말은 신용카드 거래용 단말에 비해 많이 보급되지 않았으므로, 오프라인 거래 비중이 높다면 삼성페이의 스마트폰을 활용한 IoT 기술<sup>21)</sup>이 월등히 유리하다.

21) 삼성페이 기술은 삼성이 2015년 루프페이라는 하드웨어 기반의 핀테크 업체를 인수하여 확보한 것이다. 애플 아이폰의 애플페이가 국내에 아직 들어오지 않았지만 만약 들어온다 하더라도 이런 부분은 경쟁에서 밀릴 수 밖에 없는 부분이다.



출처: 한국은행

### 3-3 개인 자산관리

전통적인 금융서비스에서 자산관리는 특정 금융기관에 계좌를 만들어 자산을 저장하고 해당 기관의 지점이나 서비스를 통해 통해서 자산을 관리하는 방식이다. 예를 들어 신한은행에 자산을 맡겼다면 신한은행 지점이나 인터넷/모바일 뱅킹을 통해 자산을 관리할 수 있다. 핀테크 기반의 자산관리는 여러 금융기관들이 오픈 API 형태로 자사의 자산관리 계좌를 개방하면, 단일 서비스를 통해 여러 계좌를 동시에 관리하는 모델이다. 사용자는 여러 금융사의 서비스를 각각 이용하지 않아도 되기 때문에 편리할 뿐만 아니라, 여러 금융사의 금융상품을 비교할 수 있는 장점도 있다. 핀테크 업체 입장에서는 여러 금융기관의 자산을 동시에 관리하기 때문에 고객의 종합적인 상황을 고려한 컨설팅 또는 상품을 추천할 수도 있다.

여기에 필요한 기술은 우선 금융기관의 자산 계좌를 개방하기 위한 오픈 API가 필요하다. 이를 기반으로 여러 계좌를 연동하는 기술과 그것을 종합적으로 저장 및 분석하는 기술이 필요하다.

### 3-4 자금이체 및 송금

자금이체나 송금은 지급결제와 매우 유사하다. 단지, 구매자를 채무자로 바꾸고 판매자를 채권자로 바꿔 생각하면 된다. 따라서 필요한 기술은 결제용 장치를 제외하고는 지급결제와 거의 유사하다. 해외 송금의 경우에는 전세계적으로 지점망을 갖추고 해외 금융기관과 제휴를 맺은 서비스 제공업체가 추가로 필요하기 때문에 이들과의 연동 방안이 필요하다.

### 3-5 디지털 화폐

디지털 화폐는 기존의 화폐를 대신하는 영역이므로 핀테크의 수준(기존 금융시스템을 편리하게 이용)을 넘어선다. 이것은 기술의 문제가 아니라 국가 수준의 화폐에 대한 지급보증 능력을 제공할 수 있는가에 있다. 화폐의 경우, 현재에도 이미 실물 화폐를 사용하는 경우보다는 전자 거래를 통해 디지털적인 숫자만 교환되는 경우가 훨씬 더 많다. 하지만 그럼에도 불구하고 디지털 화폐 대신 실물 화폐를 이용하는 이유는 국가 지불 능력이라는 벽을 넘어서지 못하기 때문이다. 이런 부분을 제외하고 순수하게 기술적인 부분에서만 본다면, 화폐의 이동(송금, 결제 등을 통해)을 기록할 수 있는 기술(기존 화폐에서는 전자결제 후 실제 금액의 차액을 맞추는 정산작업을 거치지만 디지털 화폐는 결제 자체가 화폐의 유통)이 필요하다. 현재 이런 부분을 처리할 수 있는 대표적인 기술로는 블록체인이 얘기가 되고 있다.

### 3-6 금융 업무 프로세스 개선

복잡한 금융 업무를 IT 기술로 효율화하는 부분이다. 워낙 다양한 업무 분야가 있어 한 두 가지를 특정할 수는 없다. 그리고 많은 부분이 이미 IT 시스템화되어 있기도 하다. 금융은

여러 산업 분야 중에서도 가장 많은 IT 기술이 적용되어 있는 분야 중 하나이다. 그럼에도 새로운 핀테크 기술이나 클라우드가 필요한 영역이 있다면 리스크 분석 업무가 꼽힌다. 대량의 빅데이터를 가지고 고성능 컴퓨터(HPC, High Performance Computing)를 통해 연산을 돌려 향후 발생할 수 있는 위험을 예측하는 업무이다. 이를 위해 각 금융사는 엄청난 규모의 IT 시스템을 도입하여 사용하고 있는데, 클라우드를 사용하면 이런 부분에 대한 고정비 투자를 줄일 수 있다.

### 3-7 인슈어 테크

보험은 다수의 자금을 모아 재난을 당한 사람에게 제공함으로써, 공동으로 위험을 회피하기 위한 수단이다. 문제는 재난을 당하는 사람이 많아질수록 보험료가 상승할 수 있다는 것이다. 이를 방지하기 위해 보험 가입자를 평가하여 보험료를 차등하기도 한다. 인슈어 테크는 IT 기술을 이용하여 보험 가입자 평가를 정밀하게 하는 방법이다. 예를 들어 차량 운행 기록을 IoT 기술을 통해 수집하여 위험도가 높은 운전자에게는 고가의 보험료를 적용한다. 운동을 꾸준히 하는 사람에게는 보험료를 낮게 적용한다. 또 다른 경우는 보험료를 모든 기간에 적용하지 않고 위험 상황이 발생하는 순간에만 단기적으로 적용하도록 하는 것이다. 다시 말해, 자동차 보험에서 실제 차량 운전시간을 정교하게 측정하여 해당 기간에 대해서만 보험료를 받음으로써 보험료를 차등적으로 적용할 수 있다.

### 3-8 핀테크에 적합한 클라우드 서비스

대부분의 핀테크는 기존 금융사들의 서비스들에 IT 기술을 결합하는 모델이다 보니 금융기관과의 연동이 거의 필수적이다. 이 부분에는 API 서비스가 필요하다. 은행의 핵심 코어뱅킹을 개방하기 위한 오픈뱅킹도 API 기반으로 구현이 이루어진다. 대외 연동을 위한 API 서비스는 Public 클라우드 사용이 유리하며 API 구현을 용이하게 하는 API Gateway를 직접 설치, 관리하는 것보다는 클라우드가 PaaS 형태로 많이 제공하고 있으므로 그것을 사용하는 것이 유리하다.

빅데이터 분석이나 인공지능 모델 개발도 클라우드가 제공하는 PaaS 서비스를 이용하는 것이 유리하다. 다만, 금융 데이터의 개인정보가 많이 포함된 관계로 100% Public 클라우드를 사용하기에는 규제에 걸리는 경우가 많아 Private 클라우드를 결합한 Hybrid 클라우드가 유리할 것이다. 반대로 100% Private 클라우드로 하기에는 대규모 하드웨어 구입을 하는 데 고가의 비용이 필요하고 최신 기술을 구현하는 것이 쉽지 않아 어려운 점이 있다.

빅데이터를 활용하여 고성능 컴퓨팅(HPC)을 통한 분석 작업의 경우에는 개인정보가 포함되지 않은 경우가 많아 Public 클라우드가 적합하며, PaaS 형태로 제공되는 서비스가 드물기 때문에 IaaS가 유리하다. 특히, HPC에 필요한 소프트웨어가 특수한 제품인 경우가 많아 이것도 PaaS 적용을 어렵게 하는 부분이다. 개인정보가 다수 포함된 경우라면 익명화를 통해 Public 클라우드를 활용하거나 어쩔 수 없이 Private 클라우드를 활용해야 할 것이다.

IoT 기술 또한 Public 클라우드 기반의 PaaS 서비스가 유리하기는 하지만, 핀테크 업체에서 IoT 기기와 직접 통신하기보다는 관련 업체와 연동을 하는 경우가 더 많아 해당 업체와 API 방식의 통신이 더 일반적이다.

결제 등의 거래에서 사용자를 식별하기 위한 보안 인증은 클라우드에서 해당 서비스를 직접 제공하는 경우가 드물다. 보안 인증 솔루션을 제공하는 업체와 협업을 하거나 직접 구현을 해야 하는 경우가 많아 꼭 Public 클라우드가 장점이라고 말하기는 어렵다. 이 부분은 개발, 운영에 어떤 방식이 유리할지 고민해서 결정하는 것이 맞을 듯 하다.

블록체인 기술은 Public 클라우드 업체가 PaaS 형태로 제공하기도 하지만 바로 업무에 사용할 수준을 제공하지는 못한다. 추가적인 구현 과정이 필요하다. 따라서 직접 구현하거나 기술을 가진 업체와 협업을 많이 한다. Public 클라우드에 IaaS 방식으로 구현하면 인프라 관리의 수고를 덜 수 있는 장점 정도를 고려하면 될 듯 하다.

종합적으로 판단해 보면, 핀테크 서비스를 구현할 때 빅데이터 분석, 인공지능(AI/ML), API 서비스, 고성능 컴퓨팅 등 IT 기술에 가까운 분야는 Public 클라우드가 유리하고, 구체적으로 Public 클라우드 활용은 PaaS 형태로 하는 것이 유리하다고 할 수 있다. 클라우드 기술을 통해 핀테크 서비스를 제공하고 있는 구체적인 사례는 제4절을 참고하기 바란다.

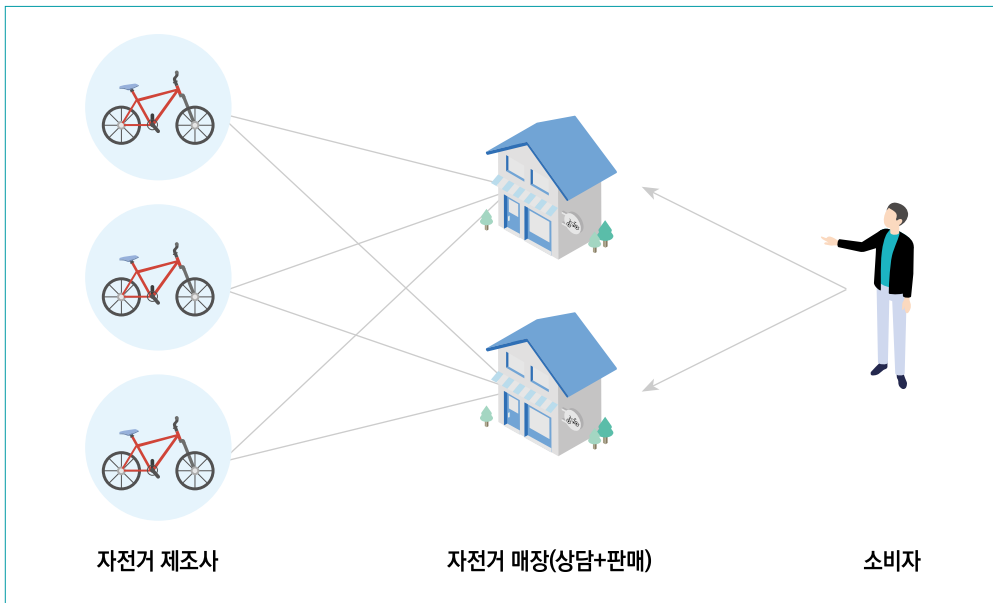
## 1 클라우드 서비스 제공자 (CSP)

여러분들이 싸이클 운동을 하기 위해 자전거를 구매한다고 가정해 보자. 여러 부품들을 모아 제조하는 자전거 제조업체가 있다. 국내에는 삼OO, 알O, 위OOO 등의 유명 업체들이 있다. 해당 브랜드의 자전거를 구매한다면, 직접 해당 업체의 직영점에서 구매하는 경우가 있지만 많은 경우 집 부근에 있는 자전거 매장에서 구매를 할 것이다. 이때, 단순히 구매 목적으로 매장에서 사는 경우도 있지만 대부분 어느 정도 싸이클 운동에 대한 정보도 듣고, 장비를 잘 활용하는 방법에 대해서도 가이드를 받는다.

싸이클을 좋아하는 매니아라면, 정기적으로 가는 매장에서 서비스와 점검도 받고 유상으로 수리를 하기도 한다. 특히 직접 브랜드 매장이나 홈페이지 등에서 구매 한다고 해서 집 부근의 매장에 비해 가격이 더 저렴하지도 않다.



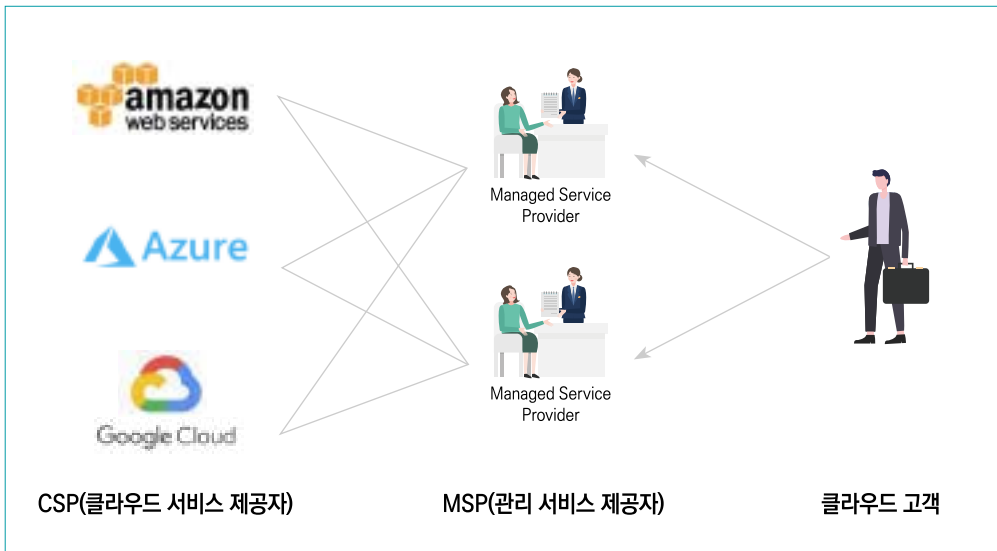
〈그림 11-12〉 자전거 비유를 통해서 보는 클라우드 공급망의 이해



직접적으로 싸이클과 비교하기는 어렵지만, 클라우드 시장에도 유사한 업체들이 존재한다. 먼저 클라우드 서비스를 구축하여 공급하는 아마존, 마이크로소프트, 구글 등의 클라우드 서비스 제공자(Cloud Service Provider, 이하 CSP)가 있고, 그들이 제공하는 각각의 서비스를 아마존 AWS(Amazon Web Services, 이하 AWS), 마이크로소프트 Azure, 구글 GCP(Google Cloud Platform)라 한다. 여러분들이 클라우드를 사용한다면 대부분은 이들 업체들이 제공하는 서비스들 중 하나를 사용한다는 의미이다. 이러한 CSP들의 서비스는 온라인상에서 회원 가입하여 쉽게 사용할 수도 있다. 예를 들어 AWS는 <http://aws.amazon.com/account/sign-up>에서 GCP는 <http://cloud.google.com/gcp>에서 직접 가입하여 사용할 수 있다. 반면, 여러분들이 클라우드 서비스에 대한 상담도 받으면서 여러 CSP들 중에서 적합한 것을 골라서 사용하고 싶다면, 이렇게 직접 가입하는 것보다 관리 서비스 제공자(MSP, Managed Service Provider)들에게 연락을 하는 것이 낫다. MSP들은 CSP로부터 클라우드 서비스를 공급받아 고객에게 제공하며, 이때 대부분의 경우 MSP는 CSP로부터 최종 가격보다 낮은 가격에 공급을 받는다. ‘규모의 경제’가 가능하기 때문이다. 따라서 MSP는 일정 부분의 마진을 고객에게 제공하는 부가적인 서비스에 투자할 수 있고,

CSP에는 부족한 서비스 및 구현 인력을 보유할 수 있다. 자전거 업체(쉬운 이해를 위해 클라우드와 비교하지만, 두 업종이 완벽히 같지는 않다는 점에 주의)는 자전거 제조에 주력을 하지만 판매나 서비스 제공은 매장의 도움을 받는다.

〈그림 11-13〉 CSP, MSP로 대표되는 클라우드 공급망의 이해



여기까지 클라우드 업계의 주요한 두 가지 플레이어인 CSP와 MSP에 대해 간략히 알아보았다. 그렇다면 클라우드 공급망을 이해하기 위해 먼저 CSP의 종류와 특징에 대해 보다 자세히 살펴보겠다. CSP의 가장 중요한 역할은 고객들이 필요로 하는 최상의 클라우드 서비스들을 구현하여 제공하는 것이다. 그러기 위해서는 고객의 목소리를 듣고 요구사항을 파악해서 자신들이 가진 기술력을 이용하여 잘 구현해야 한다. 따라서, 클라우드 사용자 입장에서는 어떤 CSP가 자신이 필요로 하는 서비스들을 잘 제공하고 있는지 파악해야 한다. 여러분들의 이해를 돕고자 주요 CSP들과 각각의 특징들에 대해 최대한 객관적인 관점에서 알아보도록 하겠다. 실제 고객이 어떤 관점에서 선택하고 있는지도 최대한 객관성을 잃지 않고 설명하도록 하겠다.

## 1-1 AWS(Amazon Web Services)

미국의 유명한 이커머스 업체인 Amazon에서 제공하는 클라우드 서비스이다. 2006년 최초로 클라우드 서비스를 시작하였으며, 사실상 클라우드 시장을 개척했다고 볼 수 있다. 참고로 개인적으로도 최초로 업무에 적용해본 클라우드 서비스이다. Amazon은 이커머스 업체의 특성상, 많은 IT 인프라를 보유 및 사용하고 있었다. 특히 특정한 기간(블랙 프라이데이)에 집중적으로 부하가 몰리는 특성으로 인해 IT 자원을 효율적으로 사용하는 방안이 많은 고민이 있었다. 따라서 이커머스에서 사용하던 IT 자원을 상대적으로 부하가 적은 시점에는 클라우드 서비스를 통해 외부에 제공할 수 있었다.

AWS의 장점은 오랜 클라우드 서비스 제공 기간을 통해 얻은 고객의 요구사항과 풍부한 경험을 바탕으로 가장 많은 서비스들을 제공하고 있다. CSP 입장에서 클라우드 서비스는 대규모 인프라 투자가 필요한 영역이기 때문에 단 시간에 모든 서비스들을 완성할 수 없다. IaaS 서비스를 시작으로 빠르게 PaaS, SaaS 영역으로 확장했고 가장 큰 시장점유율을 가지고 있다. 특히, 치열한 경쟁으로 인해 낮은 이익률을 가진 이커머스에 비해 31.3%(2016년 회계 자료 기준)의 높은 이익률을 통해 Amazon 기업 전체의 수익에 기여하고 있다.

클라우드 기반으로 IT 서비스를 구축하다 보면 기존에는 직접 설치 운영하던 소프트웨어들을 점점 관리형 PaaS 서비스로 사용할 필요성이 생긴다. 그런 면에서 AWS는 오랜 클라우드 비즈니스 경험을 바탕으로 다양한 관리형 서비스(Database, Elastic Search, LDAP)들을 제공하고 있다. 특히, 애플리케이션 개발 시 복잡한 프레임워크를 사용하지 않아도 필요한 최소한의 코드만 작성하면 바로 서비스에 사용할 수 있는 AWS Lambda와 같은 Serverless 애플리케이션 서비스의 인기가 높다. 또한 상대적으로 스타트업과 같은 소규모 기업이나 온라인 서비스 업체들이 많이 사용하던 초창기와 달리, 클라우드가 점차 IT 업계의 대세가 되면서 대규모 기업들과 전통적인 업종의 고객들이 사용하면서 소위 엔터프라이즈 고객의 요구사항들을 필요로 하게 되었다. 그러한 영역의 대표적인 사례가 바로 금융이다. AWS는 오랜 기간 클라우드 고객들을 상대하면서 이러한 부분에 대한 기능을 빠르게 개선하였다.

## 1-2 Azure

IT 소프트웨어 업계의 대표적인 기업인 MS(Microsoft)가 제공하는 클라우드 서비스이다. 2010년 서비스를 시작하였으며, 특히 2021년 현재 CEO인 사티야 나델라가 기존 기업 비즈니스 부분을 담당하다가 MS 전체의 리더가 되면서 클라우드 서비스에 집중적인 투자를 시작했다. IT 시장의 흐름이 클라우드로 옮겨갈 것이라는 그의 예상이 적중하며 비교적 빠른 시간에 클라우드 업계의 강자가 되었다.

2000년도 전후 상황을 기억해 보면, 윈도우와 오피스로 대표되는 MS의 소프트웨어를 사용하지 않는 기업이 거의 없었다. 이러한 기업용 소프트웨어 시장에서의 강점을 그대로 흡수하여 윈도우나 오피스를 사용하는 고객들을 중심으로 클라우드 서비스를 확대하였다. MS 소프트웨어를 사용하는 대기업들은 MS와 장기적인 공급계약을 하는 경우가 대부분인데, 이들에게 Azure는 쉽게 선택할 수 있는 클라우드 서비스이다. 당연히 윈도우나 MSSQL(데이터베이스의 일종) 등을 사용하는 시스템(특히 게임과 제조업계가 윈도우와 MSSQL을 사용하는 경우가 많음)을 클라우드로 전환하는 과정에는 여러모로 Azure가 장점이 되었다. 그리고 MS 특유의 편리한 UI를 적용한 PaaS 서비스와 넓은 개발자 커뮤니티(MS Visual Studio는 과거 대표적인 개발툴 중의 하나)를 중심으로 한 개발 도구들이 Azure 서비스의 특징이기도 하다.

사티야 나델라 CEO 취임 이후 눈에 띄는 눈에 띄는 변화 중의 하나는 파트너십과 개방성이다. 그 이전까지 MS는 대표적인 반오픈소스 진영으로 알려져 있었지만, 최근에는 대표적인 오픈소스 운영체제인 Linux를 적극적으로 서비스에 채택하고 있다. 그리고 기업용 데이터베이스 서비스 강자인 오라클과도 파트너십을 통해 공동으로 클라우드 서비스를 제공하여 업계를 놀라게 했다(MS의 기업용 데이터베이스 MSSQL은 오라클 데이터베이스의 강력한 경쟁 상대). 이러한 파트너십과 개방적인 정책인 기존의 기업 이미지를 긍정적으로 변화시켰고 실질적으로 고객들에게도 편의성을 제공할 수 있었다.

### 1-3 GCP(Google Cloud Platform)

검색 서비스와 유튜브로 유명한 인터넷 소프트웨어 구글(Google)이 제공하는 클라우드 서비스이다. 공식적으로는 2008년 App Engine이라고 하는 PaaS 서비스를 통해 서비스를 시작하였다. 검색이나 유튜브, 지메일 서비스 등은 누가 보더라도 IT 인프라가 많이 필요한 서비스이다. 구글은 전 세계적으로 10억 명 이상의 사용자를 가진 서비스가 7~8개에 이를 정도로 거의 전 세계 톱 IT 서비스 제공 업체이다. 그러한 대규모 IT 인프라 시스템과 운영 경험을 바탕으로 성공적으로 클라우드 서비스를 확대해 나갔다.

GCP의 특징은 먼저 구글이라는 기업이 가진 특성인 데이터 분석과 AI/ML을 바탕으로 한 서비스이다. 구글은 전세계적으로 가장 많은 데이터를 보유하고 처리하는 기업이다. 그러다 보니 해당 노하우가 들어간 Google BigQuery 데이터 분석 서비스는 많은 기업들이 대량 데이터 처리에 활용되고 있다. 또한 데이터 분석의 다음 단계 또는 최종 목적이라 할 수 있는 AI/ML 분야에서도 구글 AI/ML 기술을 기반으로 한 서비스들을 제공하고 있어 장점이 되고 있다. 최근에는 AI/ML을 쉽게 할 수 있는 AutoML이라는 기술을 적극적으로 활용하여 초보자들도 쉽게 AI/ML을 구현할 수 있게 하고 있다. 다음으로는 구글의 대규모 글로벌 서비스의 기반 인프라를 활용한 네트워크 등의 IaaS 서비스와, 오랜 오픈소스 노하우를 바탕으로 한 PaaS 애플리케이션 개발 및 운영 플랫폼인 Kubernetes(오픈소스 기술 중 하나) 등이 장점이다.

구글의 많은 서비스(검색, 유튜브, 안드로이드)들은 광고, 마케팅 그리고 모바일 서비스에 기반하고 있다. 따라서 직접적으로 클라우드의 서비스는 아니지만 그에 관련한 서비스들이 많다. 예를 들어 웹 사이트를 운영하고 있는 많은 기업들이 마케팅 성과 분석에 사용하고 있는 구글 애널리틱스(<http://analytics.google.com>)라는 서비스가 있다. GCP는 이러한 광고, 마케팅 성과 분석 도구들과 연동할 수 있는 장점을 제공한다. 따라서 고객 데이터를 분석해서 광고나 마케팅 또는 모바일 서비스(Firebase라고 하는 모바일 플랫폼 서비스도 널리 사용되고 있음)에 활용하고자 하는 경우에 유리하다.

## 1-4 국내 및 기타 클라우드 업체

여기서 설명한 AWS, Azure, GCP 이외에도 베어메탈(가상 머신 대신 물리적인 하드웨어를 통해 고성능 제공)에 강점이 있는 IBM Cloud(2011년 시작, 2013년 SoftLayer 인수), 오라클 데이터베이스에 강점이 있는 OCI(Oracle Cloud Infrastructure, 2016년 시작), 중국 서비스에 유리한 Ali Cloud(2009년 이커머스 업체인 알리바바에서 시작) 등이 있다. 이러한 업체들은 모두 해외 글로벌 업체들이다.

클라우드의 국가의 장래를 결정할 수 있는 주요 기간 산업이다. 따라서 애국심에서만 아니라 적절한 경쟁을 바탕으로 한 건강한 생태계 관점에서 국내 클라우드 서비스들의 성장도 중요하다. 마치 식량 산업처럼 당장은 비용에서 유리한 수입산을 사용하는 것이 좋지만 전시에 안보위협이 되는 것처럼, 국가의 주요 시스템이 모두 경제적인 관점만 고려하여 결정되는 것은 올바른 접근은 아니다(필자의 개인적인 견해를 밝힘). 다만, 대규모 인프라 투자가 필요한 클라우드 산업의 특성상 국내 시장만을 대상으로 한 클라우드 서비스는 경쟁 면에서 불리할 수밖에 없다. 따라서 국내 클라우드 업체들도 국내 시장만이 아닌 해외 서비스를 통해 규모의 경제를 달성하기 위해 노력하고 있다. 건강한 생태계의 발전과 국가 경쟁력을 위해서 국내 클라우드 업체들의 성장도 기대해 본다.

가장 앞서 있는 국내 업체는 네이버로 NCP(Naver Cloud Platform) 클라우드 서비스를 2017년부터 제공하고 있다. 빠르게 약 150개의 서비스(NCP 홈페이지 기준)들을 확충하여 글로벌 CSP들과 어깨를 겨루고 있다. 국내 CSP답게 우리나라의 규정 준수가 필요한 공공과 금융 분야에서 강점을 보이고 있다. 특히, 공공과 금융 분야를 위해서는 다른 사용자들과 자원을 공유하지 않는 별도의 시스템을 만들어 각종 국내 인증을 통과했다. 한국어에 특화된 AI/ML에서도 일부 강점을 보이고 있다. 해외에서는 미국 서부, 일본, 홍콩, 싱가포르, 유럽 등에 데이터센터를 두고 서비스를 제공하고 있다.

그 외에는 KT u-Cloud 서비스와 현재 서비스를 준비 중인 Kakao i-Cloud(2021년 6월 현재 홈페이지에서 보이기 는 하지만 일부 고객 대상 제한적 액세스만 가능) 등이 있다.

## 1-5 CSP 선택 기준

마지막으로 앞에서 언급한 내용과 상충이 되기는 하지만, 여러분들의 클라우드 선택이 반드시 건전한 생태계 관점 즉, 거시적일 필요는 없다는 얘기를 하고 싶다. 여러분들이 핀테크를 위해 CSP를 선택해야 한다면 그것은 여러분들의 여건과 서비스 목적에 맞는 서비스인지 여부가 가장 중요하다. 목적에 부합하는 CSP가 거시적인 관점에서도 도움이 된다면 금상첨화일 것이다. 다행인 것은 IaaS 같은 경우는 기술이 많이 평준화되어 CSP들 사이에서 차이점이 많이 사라졌다. 따라서 IaaS 위주의 서비스라면 CSP 선택의 폭이 넓을 것이다.

## 2 클라우드 관리 서비스 (MSP)

앞서 클라우드 업계의 플레이어들에 대해 설명하면서 자전거의 예를 들었고, MSP (Managed Service Provider)는 자전거 매장에 비유했다. 다소 단순화된 비교라는 생각은 들지만 실생활 관점에서 본다면 MSP이 중요한 역할을 한다는 점과 여러분들이 실제로 클라우드 도입 시 MSP가 여러분들의 곁에 있을 것이라는 점을 강조하고 싶었다. CSP가 제공하는 클라우드 서비스를 직접 구매한다고 해서 못할 것은 없지만, 정말로 클라우드를 잘 활용하고 싶다면 MSP들이 제공하는 노하우가 큰 도움이 될 것이다. 그렇다면 MSP가 제공하는 가치(value)는 무엇일까? 클라우드 도입 과정별로 MSP의 역할에 대해 살펴보자.

### 2-1 비즈니스 전략 수립

클라우드 도입은 단순히 특정 목적의 IT 시스템을 구매하는 것과는 다르다. 물론, 클라우드 서비스를 특정 목적에 따라 일부 도입할 수도 있고 그런 경우는 예외일 수도 있다. 하지만, 장기적인 관점에서 기업의 여러 부문에 클라우드를 사용해야 한다면 비즈니스 전략을 수립하는 것이 맞다. 왜 클라우드를 도입해야 하며 어떤 목적(이점)을 달성하고자 하는지 판단해야 할 것이다. 그리고 나서 어떤 부문에 언제, 어떤 범위로 도입할지 결정해야 한다. 여기서 IaaS, PaaS, SaaS와 Public, Private, Hybrid 중에서 선택을 하게 된다.

이 단계에서 MSP 역할은 신뢰할 수 있는 조언자 즉, Trusted Advisor이다. 모든 일은 첫 단추를 잘 꿰어야 한다. MSP는 고객과의 대화를 통해 고객의 비즈니스 요건이 무엇인지 파악한다. 그러면 MSP는 다양한 CSP 서비스들을 알고 있기 때문에 어떤 서비스가 고객의 비즈니스 요건과 부합할지 조언을 제공할 수 있다. 여기서 조언은 단지 하나의 정답이 아닐 것이다. 고객 입장에서도 여러 가지 판단(예산, 기술 등) 있기 때문에 그에 맞는 선택지를 제공하면, 고객은 상황에 맞게 판단을 내릴 수 있다. 여기서 필요한 MSP의 역량은 CSP 서비스에 대한 이해도 있지만, 고객의 비즈니스 요건을 정확하게 추출하는 것이 중요하다.

이러한 비즈니스 전략 수립 과정에서의 MSP 컨설팅은 반드시 유료서비스, 또는 반드시 무료서비스가 아니다. 다시 한 번 자전거의 예로 돌아가 보자. 자전거를 선택하기 위해 매장에서 본인의 필요성(로드, MTB)과 본인의 역량을 전달하면, 매장에서는 적당한 모델과 잘 사용할 수 있는 방법을 제시한다. 통상 이런 과정에 비용을 지불하지는 않는다. MSP 입장에서는 좋은 상담을 제공했을 때, 고객의 신뢰를 확보하고 매출을 올릴 수 있기 때문이다. 그러나 예를 들어 로드 사이클링을 하고 싶어하는 고객이 해당 분야 자체를 몰라 정식으로 교육과정(국내에서 드물긴 하지만 사이클링 관련 교육을 하는 기관이 있음)을 거쳐야 한다면, 그것은 무료로 제공하기 쉽지 않다. 물론 매장에서 그것까지도 무상 서비스로 제공할 수도 있지만, 그것은 어디까지나 마케팅 프로모션 관점이다. 굳이 이렇게까지 비유를 통해 자세히 설명하는 이유는 IT 현장에서 일하고 있는 입장에서 실질적인 얘기를 전달하고 싶어서이다. 이러한 사전 컨설팅 또는 거창하게 컨설팅까진 아니어도 의사결정에 필요한 사전 전략 수립 과정은 너무나 중요하다. 여러분들이 클라우드 도입에 관련한 컨설팅을 받아야 한다면 MSP와 투명하게 어느 정도까지 무료로 가능한 부분이며, 어떤 부분부터는 유료로 해야 하는지 커뮤니케이션해서 필요한 부분을 선택하는 것도 좋은 전략이다. 서로 동상이몽을 하는 경우가 말로 최악의 상황이다. 오픈 커뮤니케이션을 통해서 서로 기대치를 맞추고 가면 보다 좋은 결과를 얻을 수 있다.



### TIP

좋은 MSP를 선택하는 방법을 가이드 한다면, 두 가지 상황에서 고려해 보기 바란다. CSP는 서비스 자체의 기능이 가진 특징점이 선택 기준이지만 MSP라면 기술역량과 커뮤니케이션이 중요한 부분이다.

#### IT 기술에 익숙한 상황

여러분 스스로가 IT 기술에 익숙한 경우라면 기술 역량 자체를 중점적으로 보되 MSP의 구현 사례를 참고로 해서 Risk 부분을 어떻게 회피(hedging)하는지 보면 좋을 듯 하다. MSP로부터는 기술 역량과 경험에서 도움을 받고 실제 리딩은 직접 하면 된다. 단, 여러분들이 이미 클라우드에 익숙하여 CSP가 제공하는 기본적인 기술지원 서비스(MSP의 기술 경험 제외)가 충분하다면 MSP를 통한 커뮤니케이션 오버헤드가 발생할 수 있으므로, 주요한 메시지는 MSP와 함께 CSP에게도 동시에 전달하는 것이 효과적일 수 있다.

#### IT 기술에 익숙치 않은 상황

문제는 여러분이 IT 기술에 익숙치 않은 경우인데, 이럴 때는 기술 역량 판단이 거의 불가능에 가깝다. 전문 용어가 조금만 나오면 머리가 복잡해진다. 필자의 IT 현장 경험을 바탕으로 조언을 한다면, 아무리 어려운 IT 개념이라도 수준에 맞추어 쉽게 설명하는 MSP를 선택하라고 하고 싶다. 경험이 많고 역량이 높은 사람일수록 어려운 용어도 상대방에 맞춰 설명할 능력이 있다. 그리고 높은 공감 능력을 가진 사람이라면 상대방을 위해 쉽게 설명한다. 즉, 좋은 역량을 가졌으면서도 대화가 되는 파트너를 만날 수 있을 것이다.

## 2-2 도입 및 구현

전략 수립 과정에 도입에 필요한 준비를 마쳤다면 다음은 이행 즉, 실제 이전(migration)이나 구축(implementation)을 해야 하는 과정이다. 앞에서 결정된 시스템들을 대상으로 CSP가 제공하는 IaaS, PaaS, SaaS 중에서 골라 이전이나 구축을 한다.

이 단계에서 MSP의 역할은 기술 파트너 또는 구축 당사자이다. 고객에게 자체 인력이 있고 스스로 클라우드 구축 작업을 진행한다면 MSP는 해당 과정에서 필요한 경험이나 기술지원 등을 제공한다. 고객이 자체 인력이 있어도 전문 구축 업체(SI, System Integrator)를

아웃소싱하기도 한다. 업무는 고객이나 SI에서 담당하고 MSP는 클라우드 서비스에 대한 기술지원을 수행한다. 사실 이 부분에서 경계가 약간 모호하기는 하다. IT 서비스는 업무를 IT 기술로 구현한 것이므로 어디까지가 업무이고 어디까지가 IT 기술 즉, 클라우드의 영역인지 애매할 수 있다. 그래서 MSP의 고객 업무 이해 능력이 중요한 이유이다. 하지만, 전체 구축이 아닌 기술 파트너 역할만 하는 경우에는 MSP가 100% 업무를 이해할 것이라는 기대는 현실적으로 어렵다고 본다.

때로는 MSP가 전문 구축 역량까지 있는 경우 혹은 SI 업체가 MSP 역할까지 동시에 하는 경우에는 전체 구축을 MSP가 맡아 진행하기도 한다. 그러면 업무와 클라우드 IT 기술 사이의 경계가 애매해도 MSP 자체적으로 해결할 수 있는 장점이 있다. 다만, 고객 입장에서는 구축 비용까지 지불해야 하므로 부담이 될 수 있다.

아무리 좋은 자전거도 잘 못타면 소용이 없다. 클라우드 서비스도 활용 역량에 따라 효과가 달라진다. 기술적으로 업무 시스템과 맞지 않는 경우도 있다. MSP의 역할은 여기서 필요하다. 클라우드 서비스를 고객 여건에 맞게 잘 전달하는 것이 중요하다. MSP는 고객과 CSP를 연결하는 다리 역할을 한다. 간혹 CSP가 직접 이런 서비스를 제공하는 경우가 있기는 하지만, MSP 없이 단독으로 하기에는 CSP의 기술인력 리소스 측면에서 한계가 있는 경우가 많다.

## 2-3 운영 및 관리

클라우드 도입 및 구축이 끝났다면 이제는 운영과 관리가 필요하다. 클라우드 구축이 끝났다는 것은 이제 클라우드 기반의 업무 시스템을 본격적으로 사용하기 시작한다는 의미이다. 따라서 시스템이 잘 운영되는지 안전성을 관리하고, 비용이나 기능 면에서 최적화가 필요한지 검토해서 개선도 해야 한다. 장기적으로 새로운 서비스의 도입도 필요하다면 고민을 시작해야 하는 시기이다.

사실 MSP 역할의 핵심은 이 과정이다. 앞의 과정들을 통해 MSP가 각각 Trusted

Advisor, 기술파트너 역할을 수행했다면 이제는 한 가지에 집중한 것이 아닌 여러 가지 역량이 복합적으로 필요한 시기이다. 시스템의 안정적인 운영을 위해서는 기술적인 역량이 중요하지만, 서비스의 최적화나 신규 서비스 도입(Bigdata, AI/ML)을 위해서는 고객의 요구를 잘 파악해서 적절한 가이드를 제공하는 Trusted Advisor 역할이 필요하다. 이 과정에서 MSP는 고객과의 합의에 의해 몇 가지 모델 중에서 적절한 방식으로 서비스를 제공한다.

#### • 단순 기술지원

- 클라우드 시스템에 대한 운영은 고객이 직접하고, 기술 문제가 있을 때 필요한 지원만 제공
- 고객에게 자체 인력이 있고 비용을 최소화할 경우

#### • 운영 대행

- 클라우드 시스템에 대한 운영을 MSP가 대행. 클라우드 운영 콘솔 자체를 MSP에서 관리
- 고객이 일정 비용(클라우드 서비스 비용과 별도)을 지불하고 운영(모니터링, 구성 변경)을 위임

#### • 최적화 컨설팅

- 클라우드 시스템 운영을 넘어 리아키텍처링을 통해 비용 최적화 또는 서비스 성능 개선 도모
- MSP가 장기적 운영 과정에서 도출된 리소스 효율화(비용 절감), 신규 서비스 도입 등을 적극적으로 시도

이 부분은 일반적인 관점에서의 분류이지 모든 MSP가 동일한 서비스 모델을 가지고 있다는 의미는 아니다.

### 3

## 클라우드 도입 과정 (클라우드 도입 전략 및 고려사항)

클라우드를 도입하는 과정은 대략 다음과 같다. 각 과정에 대해서 알아보고 어떤 관점에서 접근하면 좋을지 알아보자.

- 전략 수립
- 요구 분석 및 아키텍처 설계
- 도입 및 구축
- 운영
- 개선

### 3-1 전략 수립

클라우드 도입을 위한 전략 수립 과정이다. 왜 클라우드를 도입하는지 목표를 명확히 할 필요가 있다. 즉, 클라우드가 제공하는 장점 중에서 조직이 원하는 가치는 무엇인지 매핑을 해서 각각 어떤 비중을 부여할지 결정한다. 클라우드의 장점과 조직이 가져갈 수 있는 가치들은 다음과 같다.

- 민첩성(agility): 클라우드가 제공하는 가장 중요한 가치. 낮은 초기 비용 + 운영 효율성 + 협업 능력 향상으로 인해 의사결정 시간 단축
- 비용 절감 및 고정비용 지출 축소: 신규 서비스를 작게 시작 가능. 대규모 고정비용(capex) 투자를 하지 않고 운영비용(opex)으로 지불 가능
- 운영 효율: 높은 확장성을 통해 사용자 증가로 인한 운영 위험성 감소. 표준화로 인력 수급 용이. 관리/모니터링을 위한 다양한 도구 제공
- 협업 강화: 표준화된 시스템을 통해 커뮤니케이션 부하 감소. 조직의 문화 개선 가능
- 비즈니스 역량 강화: 빅데이터 분석, 인공지능(AI/ML) 등 최신 기술을 통해 비즈니스 및 서비스 품질 향상

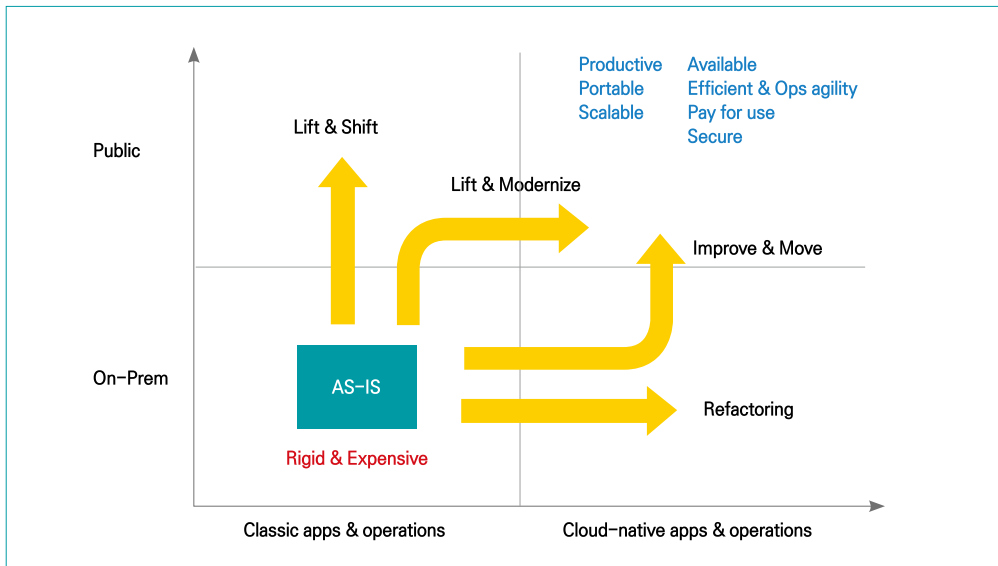
개선은 항상 현재 상태 파악에서 시작된다. 위에서 제시한 비즈니스 가치가 얼마나 필요한지 파악하기 위해 현재 상태를 파악하기 위한 평가(assess)를 진행할 수도 있다. 기존 IT 환경에 익숙한 조직에 클라우드 도입은 혁신(digital transformation)을 가져올 수 있는 좋은 방법이며, 이를 위해서는 조직 구조의 변경도 염두에 두어야 한다.

### 3-2 요구 분석 및 아키텍처 설계

클라우드를 도입하기 위한 기술 아키텍처를 수립한다. Public, Private 또는 Hybrid 클라우드 중에서 현재의 조직과 비즈니스 목표에 적합한 것을 선택한다. 기존의 시스템을 클라우드로 마이그레이션해야 하는 상황이면 몇 가지 전략 중에서 선택을 해야 한다.

- Lift & Shift: 기존 OnPrem에 있는 애플리케이션을 구조적 개선 없이 그대로 클라우드로 들어서 옮기는 방식. 처음에 클라우드로 전환할 때, 가장 쉽고 빠르게 가는 방법
- Lift & Modernize: 기존 OnPrem에 있는 애플리케이션을 먼저 클라우드로 옮긴 후에 클라우드에 적합한 형태로 개선
- Refactoring: 기존 애플리케이션 먼저 클라우드에 적합한 형태로 개선
- Improve & Move: 기존 애플리케이션 먼저 클라우드에 적합한 형태로 개선 후 클라우드로 이전

〈그림 11-14〉 클라우드 마이그레이션 전략



클라우드를 신규로 구축하던 기존 시스템을 마이그레이션하던 새로운 클라우드 환경에 대한 기반 시스템(foundation)을 잘 설계해서 구축할 필요가 있다. 클라우드는 한 번 도입되면 지속적으로 운영되어야 할 시스템이고, 전사적으로 사용되어야 하므로 기반 아키텍처를 잘 잡아 놓을 필요가 있다. 이것을 클라우드 파운데이션이라 한다.

### 3-3 도입 및 구축

클라우드 공급자를 선정하고 실제 구축작업을 진행하는 과정이다. 업무별로 IaaS, PaaS, SaaS 중에서 적절한 모델을 선정하고 상세 설계를 거친다. 예상되는 부하에 맞게 사이징을 결정하되 오토스케일링 기능을 통해 유연하게 확장 가능하도록 설정한다. 마이그레이션인 경우에는 온프레임의 데이터를 옮기는 작업도 병행되어야 한다. 데이터가 많을수록 시간이 많이 걸리고 시행 착오도 발생하므로 충분한 시간을 잡고 예행연습을 통해서 진행한다. 다운타임을 최소화하기 위해서는 전체 데이터를 옮긴 후 추가적으로 발생하는 변경분을 옮길 수 있는 방안을 마련해야 한다. 마이그레이션이 완료되면 최종적으로 기존 서비스를 중단하고 엔드유저가 클라우드에 있는 신규 서비스로 접속할 수 있도록 endpoint를 변경한다.

### 3-4 운영

클라우드 구축이 완료되면 운영 모드로 전환한다. 운영 과정에서는 다음과 같은 부분들을 고려해야 한다.

- **프로비저닝 자동화 및 오케스트레이션:** 클라우드는 모든 것을 커맨드나 API로 제어가 가능하다. 신규 자원을 할당 받는 작업도 자동화를 해서 매뉴얼한 작업을 최소화해서 실수 가능성을 줄인다.
- **모니터링 및 미터링:** 클라우드는 모니터링 시스템이 잘되어 있다. 그것을 이용하여 자원의 이상 유무는 물론 자원이 효율적으로 사용되고 있는지 관찰한다.
- **거버넌스와 정책:** 클라우드는 자원을 쉽게 만들 수 있는 만큼 무분별하게 사용되어 비용 낭비나 보안 이슈가 발생할 수 있다. 자원 생성이나 운용이 일관된 거버넌스 정책에 이루어질 수 있어야 한다.
- **비용 관리 및 최적화:** 모니터링을 통해 자원들이 효율적으로 관찰하기도 하고 비용 리포트와 비교하여 비용 최적화 작업을 지속적으로 해야 한다. 클라우드는 이를 위해 여러가지 도구들을 제공한다.
- **용량 및 자원 최적화:** 비용 최적화와 연관되는 내용이다. 모니터링을 통해 사용량이 작은 리소스가 있다면 줄이거나 다른 것과 통합하여 최적화 과정을 수행한다.
- **SLA 관리:** 클라우드 모니터링 기능에는 대부분 SLA 관리 기능을 제공한다. 서비스 응답 시간 등이 요구사항대로 나오는지 관리하여 미달할 경우 개선 작업을 수행해야 한다.
- **보안:** 클라우드에서는 전체적인 보안 거버넌스를 위해 종합적으로 점검하는 도구를 제공한다. 그런 것들을 이용하여 지속적으로 보안 상황을 체크한다.
- **지속적인 운영 및 구성 자동화:** 클라우드 운영을 효율화 하는 방법은 매뉴얼한 방법을 줄이고 자동화를 하는 것이다. 애플리케이션 배포와 같은 운영 작업을 최대한 자동화하도록 구성한다.

### 3-5 개선

운영 과정에서 나온 비효율적인 부분을 개선하는 과정이다. 빅뱅 식의 변화보다는 점진적인 개선을 추천한다.

## 1 우리가 넷플릭스 영화를 쉽게 볼 수 있는 이유 (미디어 사례)

### 1-1 클라우드 이전 배경

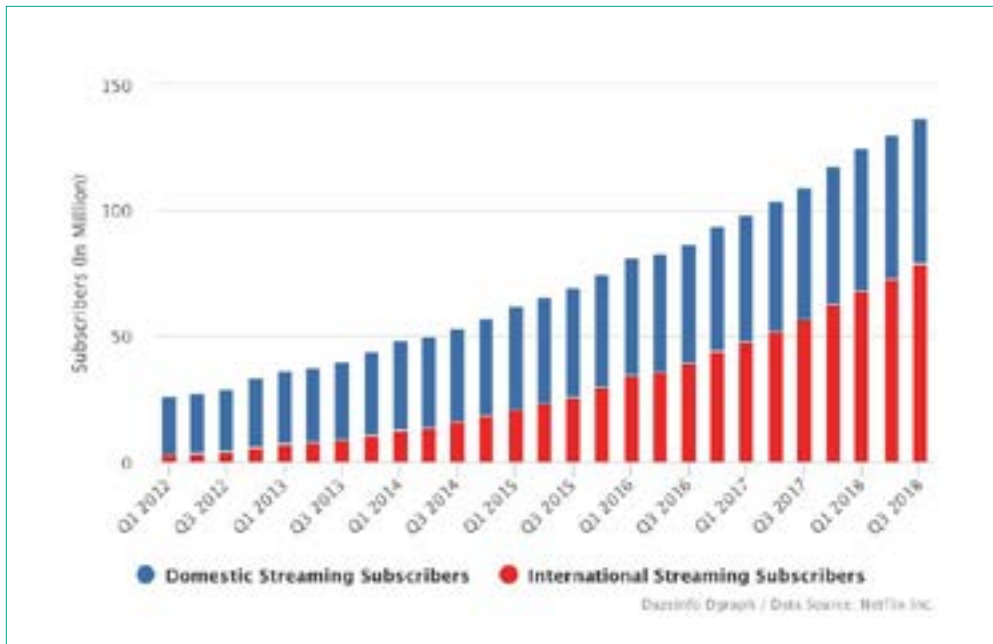
2016년 미국 라스베이거스에서 열린 CES 행사에서 넷플릭스 CEO인 리드 헤이스팅스는 한국을 포함한 130여 개 국가에 신규 진출한다고 발표했다. 그 발표가 진행되는 중에 실제로 넷플릭스 홈페이지에서는 한국 서비스가 정식으로 올라왔다. 그리고 그와 비슷한 시기에 아마존의 클라우드 서비스인 AWS가 한국에 데이터센터(클라우드 업계에서는 특정 국가나 도시에 만든다는 의미에서 통상 리전이라고도 함)를 오픈했다. 두 글로벌 서비스가 한국에 비슷한 시점에 진출한 것은 과연 우연일까?

넷플릭스를 모든 서비스를 클라우드 기반으로 제공하는 것으로 유명하다. 즉, 자체 데이터센터의 IT 시스템을 사용하지 않고 거의 100% 모든 서비스를 클라우드를 통해서 제공한다. 따라서 넷플릭스와 AWS가 동시에 한국에 진출한 것은 우연이 아니라 양사의 전략적인 판단에 의한 것이라 볼 수 있다. 넷플릭스의 영상 서비스는 많은 하드웨어 인프라 특히 컴퓨팅 자원과 네트워크, 스토리지 등을 필요로 한다. 그리고 그것을 해외에서 국제 네트워크 망을 통해서 서비스하기에는 충분한 대역폭을 확보하여 높은 품질의 영상을 제공하기에는 어려움이 많다. 아마존이 한국 내 데이터센터를 통해 클라우드 기반의 IT 서비스를 제공하면, 넷플릭스는 별도의 독자적인 데이터센터를 구축할 필요 없이 AWS 서비스를 그대로 활용할 수 있다. 실제로 넷플릭스가 한국 서비스를 AWS 한국 데이터센터 오픈과 동시에 바로 런칭을 한 것을 보면 이미 사전에 어느 정도 데이터센터 운영이 되고 있었고 사전에 구축 작업을 진행했다고 예상해 볼 수 있는 대목이다.



넷플릭스는 온라인으로 영화, 드라마 등의 영상 서비스를 제공하는 업체이고 2021년 현재 전세계적으로 약 2억명이 넘는 가입자를 가지고 있는 것으로 알려져 있다. 사업 초기에 미국 내에 9000여 개 매장을 통해 비디오를 대여해주던 대기업 블록버스터에 5천만달러 인수 제안을 했으나 거절당한 이야기는 유명하다. 블록버스터는 2010년 파산을 한 반면 넷플릭스는 유료 온라인 동영상 시장에서 1위를 달리고 있다. 지난 2008년, 데이터센터의 DB가 손상되는 문제로 3일간 DVD 배송이 안 되는 문제를 겪으면서, 넷플릭스는 약 7년에 걸쳐 클라우드로의 이전을 진행하였다. 고객이 원하면 즉시 서비스를 제공해야 하는 '가용성'이 자신들의 핵심 역량인데 그것을 자체적으로 구축하고 운영하기에는 한계에 다다랐다고 판단하게 된 것이다.

〈그림 II-15〉 Number of Netflix Streaming Subscribers, by Quarter

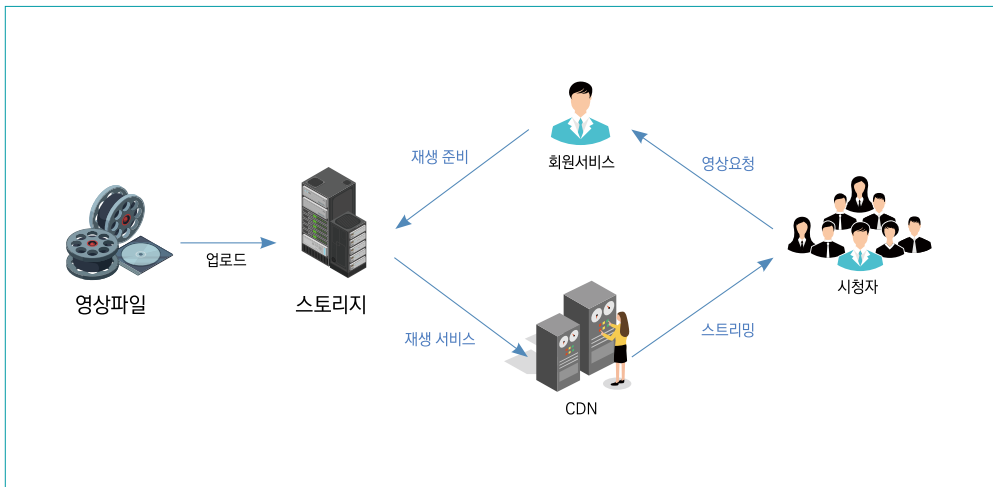


## 1-2 클라우드 이전 효과

넷플릭스가 클라우드를 선택하게 된 직접적인 동기는 ‘가용성’이었지만, 클라우드를 통해서 얻게 된 효과는 그것만이 아니다. 한국 진출 과정에서 보듯이 IT 시스템을 쉽게 늘리는 ‘확장성’을 얻을 수 있었고, 그것을 통해 원하는 비즈니스를 빠르게 제공하는 ‘민첩성’ 또한 가지게 되었다.

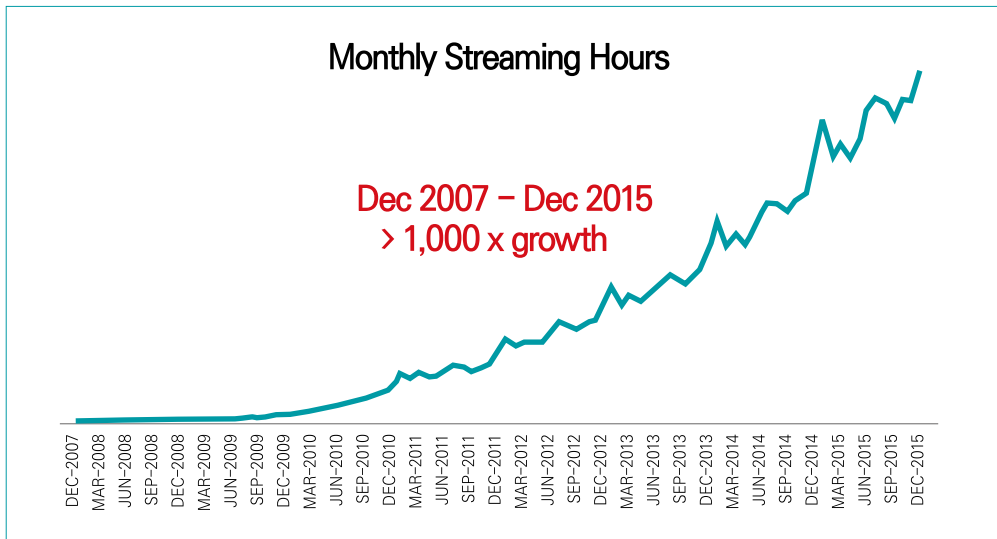
넷플릭스 같은 미디어 서비스들은 막대한 영상 콘텐츠들을 저장해 두고 그것들을 네트워크를 통해 서비스한다. 그 과정은 대략 다음 <그림 11-16> (이해를 돕기 위해 IT 시스템의 구조를 최대한 단순화 했다)과 같다. 먼저 넷플릭스와 같은 미디어 서비스에서는 영화, 드라마와 같은 영상 파일들을 확보해서 스토리지에 올려 놓는다. 영화 파일의 크기는 보통 수 GB 정도 하지만 실제로는 480p, 720p, 1080p 등 화질 등급에 따라 복수로 저장하기 때문에 더 많은 스토리지가 필요하다. 이때, 시청자가 재생 요청을 하면 가칭 ‘회원 서비스’에서 받아 회원 가입 여부, 영상 시청 정보 등을 검토한 후 스토리지에 재생 준비를 요청한다. 이 과정에 많은 컴퓨팅 자원을 필요로 하며 주로 클라우드가 제공하는 IaaS 또는 PaaS 서비스를 통해 해결한다. 마지막으로 실제로 시청자가 영상을 보는 과정은 네트워크를 통해 해당 영상 파일이 시청자의 PC나 스마트폰까지 전달되는 것으로, 통상 이러한 대규모 네트워크 기반의 콘텐츠 전송을 위해 CDN(Content Delivery Network)이라고 하는 서비스를 사용한다. 이러한 일련의 과정에 넷플릭스는 엄청난 IT 인프라를 필요로 한다. 여기에 시청자들의 회원 정보를 관리하기 위한 데이터베이스와 시청 성향을 분석하기 위한 빅데이터 서비스 또는 추천 머신러닝 서비스 등을 추가로 사용할 것이다.

〈그림 II-16〉



넷플릭스가 직접 이러한 인프라스트럭처를 직접 운영한다면 막대한 비용과 인원이 필요할 것이다. 그리고 한국 서비스의 경우에서처럼 해외 진출을 위해 각국에 데이터센터를 세우고 운영해야 한다면 추가로 막대한 비용이 소요될 것이다. 그러나 클라우드는 이러한 IT 인프라스트럭처를 단지 넷플릭스만이 아닌 다수의 고객들을 위해 이미 구축하고 제공하고 있다. 따라서 이른바 규모의 경제를 통해 저렴한 가격에 제공할 수 있고, 다수의 시스템을 운영하는 노하우를 통해 직접 운영하는 것보다 안정된 환경을 제공할 수 있다. 또한 갑작스럽게 사용자가 늘거나 줄어도 그에 맞게 탄력적으로 리소스를 사용하는 것이 가능하다. 심지어는 한국에 시스템을 운영하는 인력을 두지 않고도 글로벌 인력을 통해서 원격으로 구현하고 서비스를 하는 것도 가능하다. 넷플릭스는 이렇게 클라우드를 통해 유연성, 탄력적인 확장성 및 높은 가용성을 보장받는 것이다.

<그림 11-17>



출처: 넷플릭스 홈페이지

넷플릭스와 같은 미디어 업체들은 어떤 하나의 클라우드를 사용하기보다는 여러 개의 클라우드 업체들을 복수로 활용함으로써, 더 높은 가용성과 유연성을 얻을 수도 있다. 예를 들어 CDN의 경우 다른 클라우드 업체들이나 CDN 전문 업체들이 각자 자신들의 네트워크 인프라를 제공하고 있는데, 이들을 복수로 사용하게 되면 하나의 CDN 네트워크가 장애가 나거나 부하가 많이 몰려도 다른 업체의 서비스를 통해 이를 복구할 수 있다. 그리고 클라우드 서비스는 이러한 부분을 비교적 쉽게 연결할 수 있게 제공한다. 넷플릭스의 2015년 스트리밍 트래픽은 2008년에 비해 1천배 가량 증가했다. 이러한 확장성은 클라우드를 사용하기에 가능했던 것이다.

넷플릭스는 클라우드를 사용하면서 높은 서비스의 가용성 즉, 장애 비율도 크게 줄였다. 이러한 대규모 시스템에서 장애는 100% 피하기 어렵지만 클라우드 기반의 분산 시스템을 통해 복잡한 아키텍처 대신 단순하고 장애에 강한 복원력 강한(resilient) 시스템을 구축할 수 있었다. 비용은 넷플릭스가 클라우드를 선택하는 데 주된 동기는 아니었지만 클라우드의 규모의 경제와 고정 자산 투자 없이 필요한 서비스에만 비용을 지불하는 특징으로 인해 스트리밍 비용을 대폭 절감하였다.

### 1-3 마이그레이션 과정의 교훈

넷플릭스는 처음부터 클라우드 기반으로 구축한 것이 아니라 기존에 자체 데이터센터에서 운영하던 시스템을 클라우드로 이전(migration)한 것이다. 그리고 이전 과정은 7년이나 걸렸다. 7년이면 짧은 기간이 아니고 그 과정에 여러 가지 어려운 점들도 있을 것 같아 우리가 무엇을 배울 수 있는지 알아보려고 한다.

이전 과정에 7년이 소요된 것은 현재 서비스 중인 시스템을 다운타임 즉, 서비스 중지 없이 옮기다 보니 생긴 자연스러운 과정이라 볼 수 있다. 수십 또는 수백 가지의 시스템들을 한번에 옮기기보다는 영향도가 적은 것부터 조금씩 차례대로 이전을 했다. 반대로 처음부터 클라우드 기반으로 서비스가 구축된 것이었다면 이런 마이그레이션 과정은 불필요했을 것이다. 여기서 생각해야 할 것은 제3절에서 언급한 도입 과정과 전략에서 보듯이 기존의 서비스를 그대로 클라우드로 이전만 한 것이 아니라 클라우드에 맞게 현대화(modernization)했다는 점이다. 시스템을 운영하는 방식도 많이 개선(넷플릭스의 내부 시스템 구조는 클라우드의 장점을 극대화할 수 있는 최신 아키텍처인 마이크로서비스 기반인 것으로 유명하다)을 했을 것이다. 이를 통해서 넷플릭스는 클라우드 이전으로 얻을 수 있는 효과를 극대화했다.

반대로 만약 여러분들이 신규로 서비스를 시작한다면 처음부터 클라우드로 시작하는 것이 효과적이라 짐작해 볼 수 있다. 클라우드 기반의 시스템들은 기존 방식의 애플리케이션에 비해 향상된(확장 용이하고 가용성 높은) 품질을 제공한다. 그리고 신규로 시작하는 서비스임에도 불구하고 클라우드가 가진 IaaS, PaaS, SaaS 등을 골고루 활용하면 처음부터 기존의 대규모 서비스만큼 높은 품질의 서비스를 제공할 수 있는 것이다.

### 2-1 이커머스 산업의 이해

여러분들은 대부분 온라인 쇼핑몰에서 물건을 구매해본 경험이 있을 것이다. 그리고 오프라인 매장에서 물건을 구매해본 경험도 당연히 있을 것이다. 두 가지 경험의 공통점과 차이는 무엇일까?

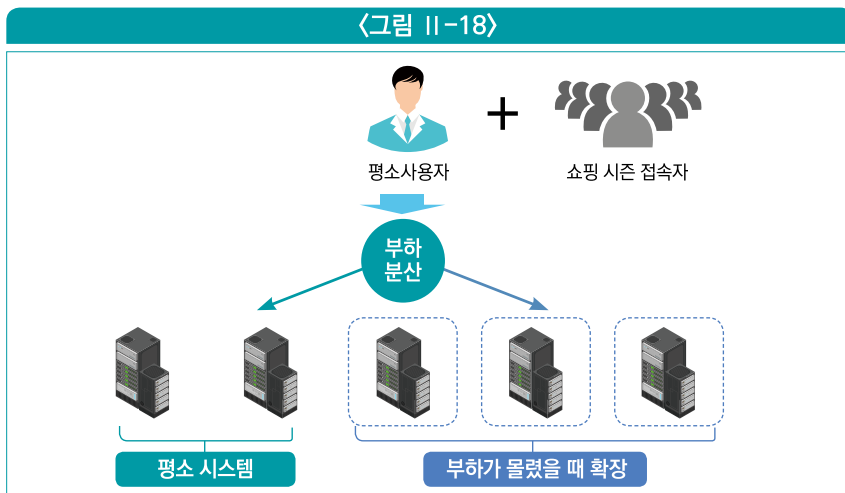
오프라인 매장에서 물건을 구매할 때 좋은 가격과 품질의 제품을 구매하기 위해 흔히 많이 했던 말이 ‘발품을 판다’라는 표현이다. 즉, 오프라인 매장은 직접 이동을 해야 하기 때문에 발로 이동하는 경우가 많다는 의미이다. 시간적으로나 공간적으로 제약이 어느 정도 생긴다는 의미이다. 하지만 직접 물건을 보고 만져볼 수도 있으므로 구매할 물건에 대해 더 정확하게 체크할 수 있다는 장점이 있다. 반면, 온라인 구매는 웹 브라우저를 통해 복수의 쇼핑몰 페이지 간에 쉽게 왔다 갔다 할 수 있으므로 시간적 또는 공간적 제약이 없다. 가격 비교도 매우 쉽다. 따라서 가격 차이가 조금 나거나 서비스의 품질(페이지 로딩 속도, 결제 방법)에 문제가 있다면 다른 온라인 쇼핑몰로 변경하는 것이 대단히 쉽다. 온라인 쇼핑 서비스를 제공하는 이커머스 업체 입장에서는 이러한 특성들로 인해 몇 가지 어려운 점들이 발생한다.

첫 번째는 경쟁이 매우 치열하다는 것이다. 사실 이커머스만이 아닌 유통 분야는 전통적으로 경쟁이 치열한 분야이다. 매출 규모에 비해 적은 이윤을 내는 경우가 많은데 대량의 상품들을 취급하지만 가격이 중요한 선택의 요소이기 때문이다. 소비자가 원하는 많은 물건을 취급해야 하기 때문에 상품 및 재고 관리에도 비용이 많이 소모된다. 이러한 유통의 특성이 온라인으로 가게 되면 몇 배는 커진다. 특히 위에서 말한 것처럼 온라인 쇼핑몰 사이에서는 변경으로 인한 시간, 공간적 제약마저 거의 없기 때문에 조금이라도 더 오래 고객을 붙잡기 위해 경쟁을 많이 하게 된다.

두 번째는 민첩성이 필요하다는 것이다. 경쟁이 치열하다 보니 조금이라도 편리한 서비스를 제공하기 위해서는 빠르게 시스템을 구축하고 안전성을 높여야 한다. 최근 코로나(COVID-19) 이후 온라인 쇼핑몰들 간에 배송 전쟁이 벌어진 것을 알고 있을 것이다.

오프라인 매장 방문이 어려워지고 특히 식품처럼 빠른 신선도가 중요한 상품의 경우에는 빠르게 배송하는 것이 너무나 중요하다. 그럼에도 배송 물량은 몇 배씩 증가하였다. 그런데 만약 1~2년 걸쳐 시스템을 구축해야 한다면 이런 중요한 기회를 놓치게 되는 결과를 가져올 것이다. 사실 코로나가 아니어도 새로운 상품이나 트렌드의 등장과 같은 변화가 많은 시장이다 보니 민첩성이 너무나 중요하다.

세 번째는 확장성이다. 우리가 쇼핑물을 자주 찾는 경우는 언제일까? 온라인이나 오프라인을 떠나 연말, 입학시즌, 명절 등에 평소보다 더 자주 찾게 될 것이다. 미국의 경우 추수감사절 다음 금요일부터 크리스마스, 새해 시즌까지의 기간을 블랙프라이데이라 해서 기업들이 그동안의 재고를 싸게 팔고, 이때 소비자들도 한 해 동안 기다렸던 지갑을 열고 많은 쇼핑을 하고 있다. 미국 연간 소비의 20%를 차지하고 일부 기업들의 경우 한 해 매출의 70%가 발생한다고 한다. 미국과 투톱을 다투는 중국의 경우 광군제, 우리나라에도 코리아 세일 페스타라는 유사한 이벤트가 있다. 이때 이커머스 업체들은 평소 시스템보다 수 배 ~ 수십 배의 시스템을 필요로 한다. 그래서 평소 사용량이 많지 않더라도 이 기간을 위해 시스템의 대규모 시스템을 유지하고 있다. 만약 평소에는 작은 규모로 시스템을 유지하고 이 시기에만 시스템을 증설할 수 있다면 효과적일 것이다. 클라우드는 사용하는 시스템의 규모를 자유롭게 늘렸다 줄였다 할 수 있고, 특히 갑자기 사용자가 몰려들 때도 자동으로 시스템을 확장하는 기능(auto-scaling)이 있어 신속하고 유연한 대응이 가능하다.



사실, 아마존이 AWS 클라우드 서비스를 출시한 이유가 바로 이러한 이커머스 산업의 특성 때문이었다. 알다시피 아마존은 대규모 이커머스 서비스를 하는 업체다. 따라서 대규모 IT 인프라를 운영했는데, 미국의 블랙프라이데이 시즌이 아닌 경우 유휴 시스템 리소스가 많았다. 그래서 그것을 외부의 다른 고객들에게 제공할 수 있도록 만든 것이 클라우드의 시작이었다. 이처럼 이커머스는 경쟁이 치열하고 매우 빠른 시장 대응이 중요한 분야이고 클라우드는 그러한 업종의 특성에 매우 적합한 서비스를 제공하고 있다.

## 2-2 이커머스 클라우드 사례

사실상 거의 모든 국내외 이커머스 업체들이 클라우드 서비스를 직간접적으로 전체 또는 일부 사용하고 있다. 이커머스 서비스를 하면서 클라우드를 쓰지 않는 것이 오히려 어려운 실정이다.

국내의 경우 대표적으로 2017년 쿠팡이 자사의 모든 시스템을 클라우드로 전환하였다. 기존에 두 군데 데이터센터를 통해 직접 시스템을 구축하고 운영을 했지만, 확장성 문제가 심각하고 발목을 잡았다. 약 1년여의 구축 및 시험 과정을 거쳐 모든 시스템을 중단 없이 클라우드로 이전하였다. 앞의 넷플릭스 사례와 비교하여 빠르게 전환이 가능했던 것은, 이미 애플리케이션의 구조가 클라우드에 적합한 마이크로서비스 아키텍처로 되어 있었기 때문이다.



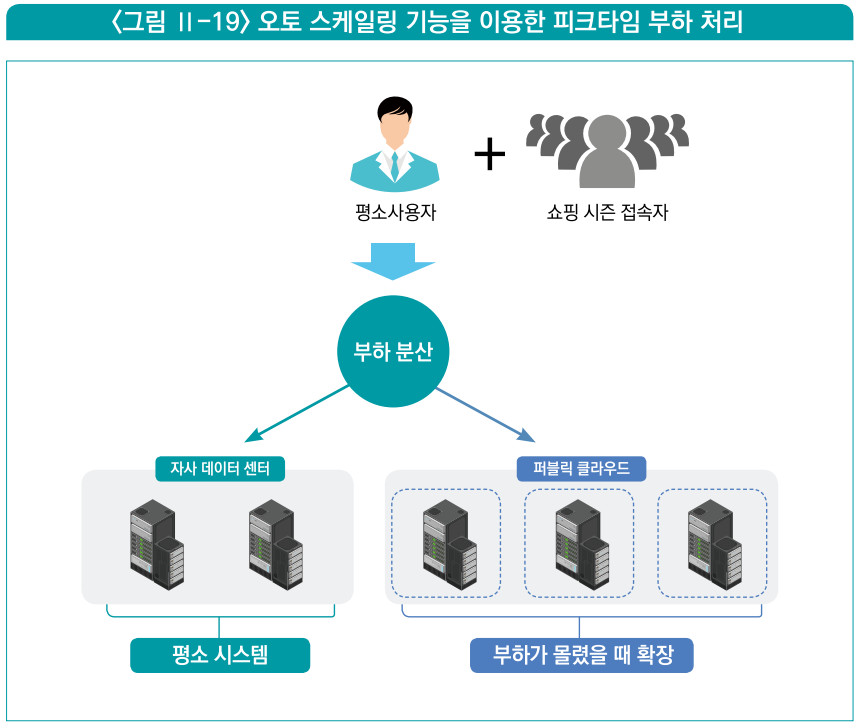
## [마이크로서비스 아키텍처]

마이크로서비스 아키텍처(Microservice Architecture, 이하 MSA)는 클라우드에 올리는 애플리케이션(응용 프로그램)에 적합한 아키텍처로 잘 알려져 있다. MSA는 기존의 애플리케이션이 하나의 큰 덩어리로 되어 있어 새로운 기능 추가를 위한 변경에 취약한 단점을 극복하기 위해, 애플리케이션을 각각 하나의 작은 서비스 단위로 나눈 형태이다. 아래 그림처럼 기존의 구조를 MSA와 대응하여 거대한 구조체인 monolith(거대한 조직 또는 구조물)라고 한다면, 작은 단위를 마치 선박에 싣는 컨테이너처럼 작은 단위로 쪼개어 관리하는 것이다. 이렇게 함으로써 부분적인 변경이 있을 때, 전체 구조(monolith)를 변경할 필요 없이 필요한 부분만 수정하여 사용할 수 있다. 만약 기존의 monolith 구조였다면 확장을 할 때도 또 하나의 큰 구조물이 필요하지만, MSA에서는 필요한 상자(container)를 쉽게 늘렸다 줄였다 할 수 있어 확장성 면에서도 매우 유리하다. 부분적으로 변경이 용이하기 때문에 새로운 서비스에 빠르게 대응하는 민첩성(agility) 면에서도 유리하다. 다만, 기존에 개발되어 있는 monolith 구조의 애플리케이션을 MSA 형태로 분해하는 것은 어느 정도의 설계 변경이 필요한 부분이라, 클라우드 전환 과정에 미리 준비가 되어 있지 않다면 이 부분을 위한 별도의 고려가 필요하다.



출처: <https://blog.knoldus.com/why-do-you-always-choose-microservices-over-me-said-the-monolithic-architecture/>

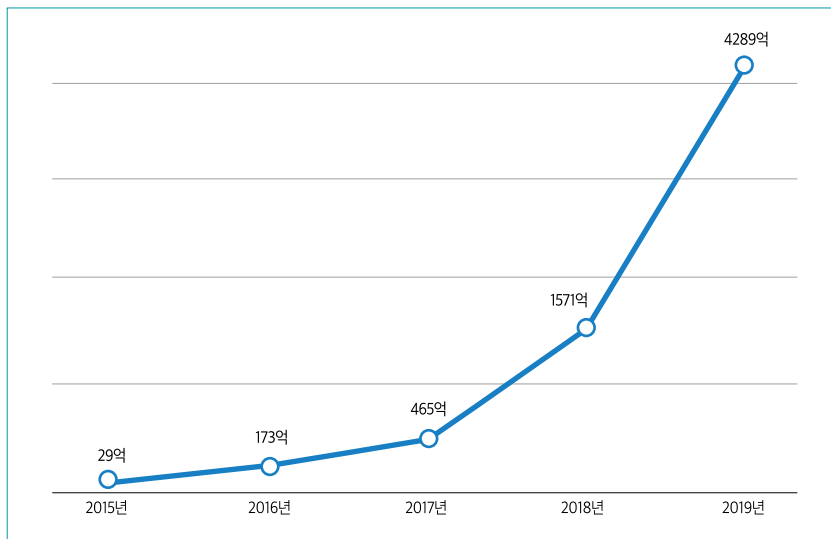
쿠팡의 서비스는 기존에는 대규모 이벤트를 진행 할 때 사이트 접속이 원활하지 않거나 장애가 발생하는 경우가 있었지만, 클라우드 전환 후에는 쉽게 시스템 용량을 증설할 수 있어 이러한 문제점을 해결하였다. 특히 새로운 기능 개선을 할 때 부분적으로 변경할 수 있어 시스템 전체에 장애가 발생할 가능성을 줄일 수 있었다. 이베이나 SSG.COM 등과 같은 국내의 다른 이커머스 업체들도 전체는 아니지만 부분적으로 클라우드를 사용함으로써 유사한 문제들을 회피하고 있다. 평소의 트래픽은 자사 데이터센터의 자원을 활용하다가 명절이나 연말과 같이 부하가 몰리는 시기에는 클라우드를 사용하는 하이브리드 형태의 구성을 활용하기도 한다.



국내의 또 다른 이커머스 클라우드 사례는 신선한 식재료를 가정에 배달해 주는 스타트업인 마켓컬리이다. 2014년 설립된 마켓컬리는 전날 주문한 식재료를 다음날 새벽에 배송하여 가정에서 바로 다음 날 아침부터 사용할 수 있게 해주는 이른바 '셋별배송' 서비스를 통해

유명해졌다. 일반 공산품에 대한 이커머스 서비스는 많았지만 신선한 식재료를 빠르게 배달해 주는 모델은 국내에서 거의 최초였고 성공적이었다. 특히 2020년 시작된 COVID-19 사태로 인해 대면식 대형마트 쇼핑에 대한 불안감 소비자들이 대거 몰리면서 매출이 1조원에 육박하였다. 마켓컬리는 이러한 이커머스 서비스를 100% 클라우드에서 운영 중이며, 이러한 사용자 폭증에 원활하게 대응할 수 있었다.

〈그림 11-20〉 마켓컬리 매출액 증가 추이



출처: 마켓컬리

쿠팡을 포함한 다른 이커머스 업체들도 경쟁적으로 유사한 새벽배송 서비스를 통한 신선식품 배송 서비스를 내놓거나 강화하여 현재 치열한 경쟁이 일어나고 있는 업종이다. 사실 신선식품은 이커머스 업체들에게는 기존의 오프라인 매장 위주의 커머스 업체들과의 경쟁에서 아픈 손가락이었다. 글로벌 이커머스 선두 업체이자 진출하는 분야마다 경쟁 업체들을 고사 직전으로 몰아붙이는 아마존도 신선식품 분야에서는 월마트에 밀리면서 고전을 하고 있는 영역이다. 아마존도 신선식품 분야의 경쟁력 강화를 위해 훌푸드마켓을 인수하여 월마트와 치열하게 경쟁 중이다.

이커머스 업체들에게 신선식품이 어려운 이유는 몇 가지가 있지만, 그중에서도 가장 어려운 부분은 바로 식품 그중에서도 유통 기간이 매우 짧은 신선식품이라는 점이다. 우선 식품을 신선하게 유지할 수 있도록 배송 기간이 짧아야 하며, 그러기 위해서는 배송 효율화가 필수적이다. 일반 공산품은 하루나 이틀 늦게 배송이 되더라도 큰 문제가 없지만, 신선식품은 상하게 되면 상품의 가치가 제로가 된다. 또한 다른 상품들과 달리 재고를 유지하는 것이 불가능하기 때문에 수요 예측이 정확히 되어야 한다. 기존 쇼핑물들의 경우, 유통 기간이 지난 신선식품 폐기율은 2~3%에 달하였다. 마켓컬리는 AI/ML 기반의 수요 예측을 통해 폐기율을 1% 미만으로 관리한다.

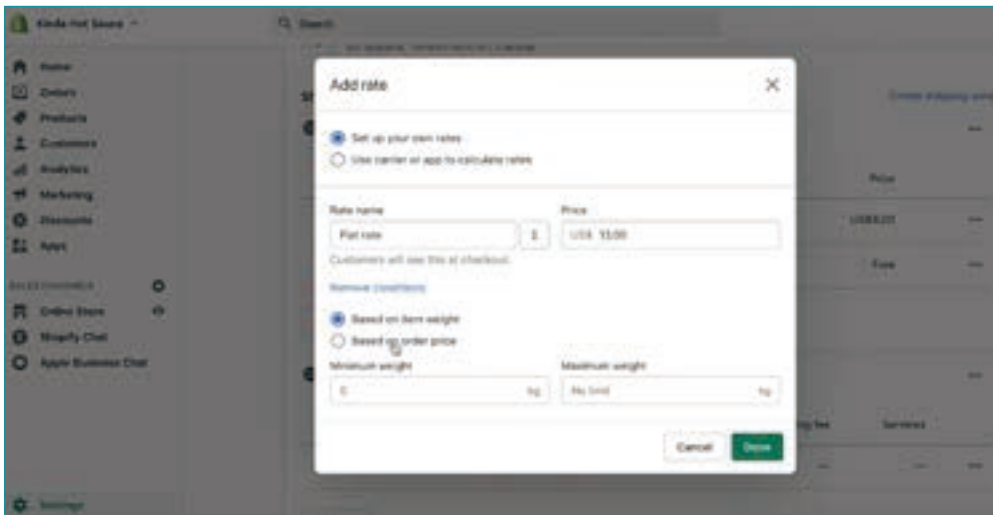
이러한 효율적인 배송과 정확한 수요 예측을 위해서는 기본적인 시스템 효율화도 중요하지만, 특히 데이터 분석과 AI/ML을 활용한 높은 정확도가 필수이다. 클라우드는 효율적이고 안정적인 시스템 운영을 가능하게 해 줄 뿐만 아니라, AI/ML 플랫폼을 제공하여 대규모 인프라에 투자를 하지 않고도 머신러닝을 업무에 적용할 수 있게 돕고 있다. 마켓컬리와 같은 스타트업 입장에서 대규모 IT 투자나 AI/ML 인프라 투자가 쉽지 않겠지만, 클라우드는 소규모 스타트업이 기존의 거대한 유통기업들과 경쟁을 가능하게 하였다.

해외로 눈을 돌려서 보면 아마존(스스로 AWS 클라우드 서비스 제공 중)과 경쟁하고 있는 글로벌 이커머스 업체인 쇼피파이(Shopify)나 영국의 신선식품 기업 오카도(Ocado) 등이 대표적인 사례이다.

쇼피파이는 많은 이커머스 업체들이 쉽게 자사의 서비스를 할 수 있도록 SaaS 기반의 이커머스 플랫폼을 제공하는 업체로 유명하다. 초기 이커머스 시장에서 페이팔(Paypal)은 중소 업체들이 결제 수단으로 많이 활용되었다. 그러나 페이팔의 구매 취소 정책의 불합리성으로 판매자들이 어려움을 겪자 자신들의 편리한 결제 수단을 제공하여 그 시장을 장악해 나갔다. 그 외에도 배송 관리, 검색엔진 최적화, SNS 지원 등의 편리한 기능을 제공하여 중소 상공업자들이 자신들의 물건을 쉽게 판매할 수 있는 플랫폼을 제공해 주고 있다. 이러한 혁신적인 서비스를 통해 현재 전 세계 이커머스 업체 중에서 아마존에 이어 두 번째로 큰 기업에 올랐다. 구글, 테슬라, GE, P&G 등 약 60만 개의 온라인 쇼핑몰이

쇼피파이 기반으로 구축되어 있다. 쇼피파이는 대부분의 시스템을 퍼블릭 클라우드 기반으로 운영하고 있으며, 스스로 자사의 이커머스 플랫폼을 SaaS 클라우드 서비스로 고객들에게 제공하고 있다. 이러한 SaaS 모델은 아마존(해외)이나 쿠팡(국내)과 같은 거대한 이커머스 업체들이 시장을 지배하는 상황에서, 기업들이 자신들만의 이커머스 서비스를 할 수 있는 환경을 제공하고 있다. 이러한 특징으로 인해 마치 이커머스 분야의 유튜브 플랫폼처럼 게임체인저로 받아들여지고 있다.

〈그림 11-21〉 이커머스 서비스를 쉽게 만들 수 있는 쇼피파이 플랫폼



출처: <https://www.youtube.com/watch?v=u-Qfdn44rB4&t=3s>

영국의 오카도는 우리나라의 마켓컬리의 롤모델이라 불릴만한 회사이다. 유통 시장에서 신선식품이 중요한 비즈니스가 될 것이라 예상하고 이 부분의 효율화에 집중적인 투자를 해서 성공을 거두었다. 그래서 신선식품 분야에서 고전을 하고 있는 아마존의 킬러라는 별명도 가지고 있다. 앞서서도 말했듯이 신선식품은 수요 예측에 기반한 공급관리와 신속한 배송이 생명이다. 오카도는 먼저 물류창고를 로봇 기반으로 완전히 자동화하여 주문부터 배송에 드는 시간을 혁신적으로 단축시켰고 자율주행트랙, 배달 경로 최적화 등을 통해 배송 효율화도 이루었다. 특히 여기에는 빅데이터에 기반한 데이터 분석과 AI/ML 기술이 들어가

있다. 오카도는 이러한 자사의 혁신적인 시스템을 플랫폼으로 만들어 다른 기업들에게도 판매하고 있다. 심지어 자사를 단지, 유통 이커머스 기업이 아니라 IT 기술회사라 말하고 있다. 오카도는 프라이빗 클라우드와 퍼블릭 클라우드를 효율적으로 결합하여 사용하고 있다. 그 중에서 오카도 플랫폼의 핵심이자 대규모 자원이 필요한 데이터 분석에는 퍼블릭 클라우드를 활용하고 있다.

〈그림 11-22〉 오카도의 자동화된 물류 창고



출처: [https://www.youtube.com/watch?v=EeMTZd68fOU&t=5s&ab\\_channel=OcadoTechnology](https://www.youtube.com/watch?v=EeMTZd68fOU&t=5s&ab_channel=OcadoTechnology)

### 3 금융/핀테크에서 클라우드 (금융/핀테크 사례)

금융산업은 한 번의 사고로도 치명적인 결과를 초래할 수 있어, 감독 당국의 엄격한 규제와 철저한 감독을 받는 분야이다. 따라서 금융(핀테크 포함)에서 클라우드가 도입되기 시작한 것은 비교적 최근의 일이다. 그럼에도 핀테크의 등장에서 보듯이 금융은 IT 기술이 매우 중요한 분야이기도 해서 앞으로 클라우드 도입은 가속화될 것으로 예상이 된다. 실제로 금융 분야에서 클라우드가 도입된 이유는 핀테크 때문인 경우가 많다. 몇 가지 클라우드 도입 사례를 살펴보자.

#### 3-1 골드만삭스 애플카드 (지급결제)

골드만삭스는 2019년 8월 자사의 첫 신용카드인 애플카드를 출시했다. 애플이 투자은행 골드만삭스와 제휴해 선보인 신용카드로, 애플의 1억 명이 넘는 미국 아이폰 고객을 기반으로 대형 투자은행인 골드만삭스에서는 필요한 금융 조달을 담당하고, 마스타카드가 전 세계에 구축된 결제망을 제공한다. 애플카드는 애플페이를 입힌 실물 카드로 티타늄 카드 형태로 제작되었으며, 일반적인 카드와 달리 카드번호 · 유효기간 · CVC코드가 없는 것이 특징이다. 결제를 할 때마다 가상의 카드번호가 일회성으로 생성되게 하여 보안을 강화했다. 지갑(wallet) 앱을 통해 신청할 수 있고, 승인이 나면 오프라인 매장에선 아이폰을 결제 단말기에 갖다 대는 애플페이를 통한 결제와 실물 카드를 이용하는 두 가지 방법으로 결제할 수 있다. 애플카드는 신용카드 관련 비용을 낮춰 캐시백 프로그램으로 결제액의 1~3%의 적립할 수 있도록 하여 고객에게 혜택을 제공한다.

〈그림 11-23〉 애플카드 모습



출처: 애플

골드만삭스는 트레이딩 부문의 매출 저하로 성장이 정체되고 있어서 소매금융에 관심을 기울이고 있다. 골드만삭스는 JP모건의 체이스 은행 같은 탄탄한 상업은행이 없기 때문에 리테일 고객 기반이 약하다. 2016년 개인과 소규모 기업을 대상으로 한 온라인 예금 및 대출 플랫폼인 마커스(Marcus)를 출시하였고, 2018년 4월에는 개인재무관리 애플리케이션을 제공하는 클래리티머니(Clarity Money)를 인수해 모바일 소매금융을 한층 강화하였다.

애플카드는 역사상 가장 빠른 성장률로 두 달만에 100억불 대출을 달성하였고, 3년 만에 550억불의 소액대출을 유치했다. 애플카드는 아마존 AWS 기반으로 구축되었으며 급증하는 트랜잭션을 성공적으로 처리하였다.

골드만삭스는 이외에도 기업들의 현금 자산 관리를 위한 Transaction Banking Service를 역시 클라우드 기반으로 출시하였다. 오픈 플랫폼으로 다양한 banking 서비스를 구현하여 수 조원의 트랜잭션을 처리하고 있다.



### 3-2 HSBC 위험도 분석 (금융 업무 프로세스 개선)

글로벌 금융사인 HSBC는 핀테크 기술을 적용하고 클라우드를 적극적으로 활용하고 있는 기업이다. 이미 오픈뱅킹 서비스를 운영 중이고, 2020년에는 금융 리스크 분석 업무를 구글 GCP 클라우드 기반으로 구축하였다. 금융당국에게 데이터 보안과 리스크 관련 답변을 신속하게 해야 하는 HSBC는 온프레미스 기반의 분석 시스템과 고성능 컴퓨팅을 활용하고 있었다. 이러한 작업은 한 번에 10시간씩 걸리는 일이었고 신속한 대응이 쉽지 않았다. 그래서 클라우드의 대규모 인프라를 활용한 데이터 분석 시스템으로 100페타바이트의 데이터를 옮겼고, 결과는 10시간 걸리던 업무를 30분만에 처리할 수 있게 되었다.

〈그림 11-24〉 HSBC 은행



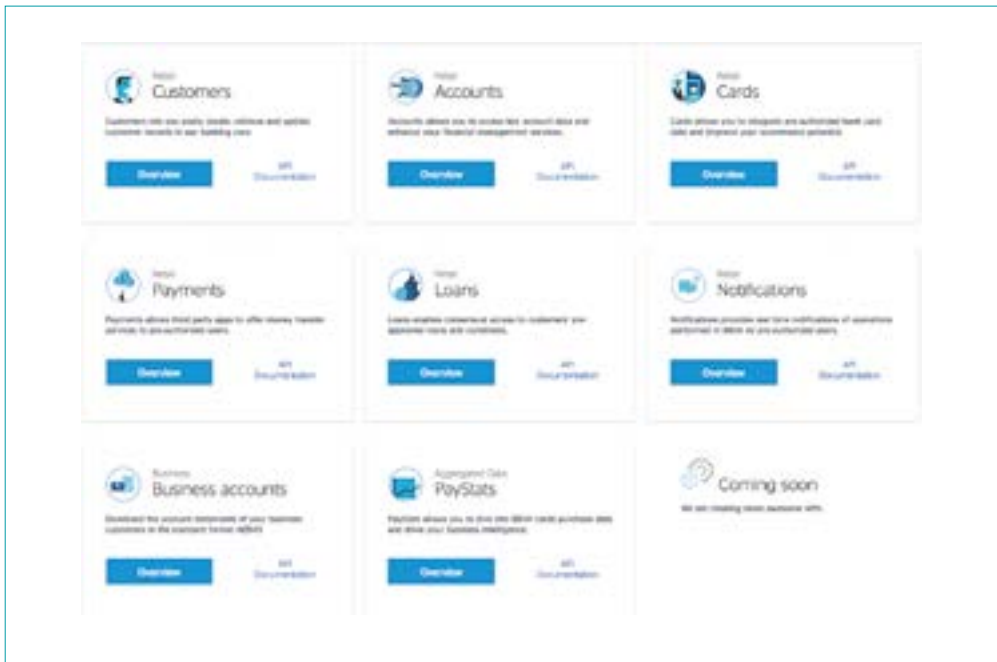
출처: HSBC

HSBC는 이뿐만 아니라 다양한 핀테크 서비스를 통해 고객의 경험을 향상시키기 위해 클라우드 우선(Cloud First) 전략을 취하고 있다. 다양한 IT 시스템을 클라우드 기반으로 전환하여 회사 내부의 문화도 혁신하고 있다.

### 3-3 BBVA 오픈뱅킹

1857년 설립된 스페인 기반의 글로벌 금융 그룹 BBVA는 전 세계적으로 8000여 개의 지점과 7400만명의 고객을 가지고 있다. 2014년에는 미국 디지털 뱅크인 “Simple”을 1700만 달러에 인수하여 핀테크 경쟁 우위를 확보하기 위한 노력을 하고 있다. 2017년과 2018년에는 2년 연속으로 Forrester 선정 글로벌 모바일 뱅킹앱 1위를 차지했다. BBVA는 API Market을 통해 자사의 다양한 금융서비스를 핀테크나 타 금융사에 오픈뱅킹 형태로 제공하고 있다. 개발자들이 쉽게 API를 사용할 수 있도록 아래 그림과 개발 포털을 제공한다. 이 서비스는 아마존 AWS 클라우드 기반으로 구축되어 탄력적으로 운영되고 있으며, 모든 개발 및 배포 과정이 완전히 자동화되어 운영된다.

〈그림 II-25〉 BBVA의 오픈 API 개발 포털



출처: BBVA

## 1 핀테크와 클라우드

핀테크는 Finance와 Technology의 합성어로 금융과 최신 IT 기술을 접목하여 새로운 금융서비스를 만드는 것을 의미한다. 핀테크 기업들은 기존 금융서비스에 클라우드, 모바일, 빅데이터 분석, 인공지능 등의 기술을 적용하여 혁신적인 서비스들을 만들고 있다. 삼성전자의 삼성 페이, 카카오의 카카오 페이, 애플의 애플 페이, 알리바바의 알리 페이 등이 있다. 핀테크의 분야는 대략 다음과 같은 것들이 있다.

- 대부 중개, 크라우드 펀딩 및 P2P 대출: 자금을 필요로 하는 사람과 자금을 투자하려는 사람 사이를 중개
- 지급결제 및 전자결제: 모바일 기기나 온라인으로 편리한 결제 솔루션을 제공
- 개인 자산관리: 여러 금융기관에 나뉘어져 있는 자산을 한 군데서 관리하고 컨설팅을 제공
- 자금이체 및 송금: 모바일 기기나 온라인으로 자금의 이체 과정을 편리하게 함
- 디지털 화폐: 기존의 화폐를 대체하는 온라인 기반의 디지털 화폐
- 금융 업무 프로세스 개선: 복잡한 금융 업무를 IT 기술을 통해 개선
- 인슈어 테크: 보험 가입자를 선정이나 보험 가입비를 계산을 IT 기술을 이용하여 정교화

전통적인 금융기업이나 IT 기업은 핀테크를 통해 여러 가지 효과를 얻을 수 있다. 첫 번째는 금융산업의 경쟁력 향상이다. 산업간의 경계가 무너지고 경쟁력이 치열해지는 상황 속에서

전통적인 금융기업들은 IT 기술을 결합해 비즈니스 영역을 확대하거나 본연의 금융서비스에 집중함으로써 경쟁력을 높일 수 있다. 특히 금융과 IT 기술이 융합되면서 혁신적인 서비스가 탄생함으로써 새로운 수익 모델이 만들어지고 시장 자체가 성장할 수 있다. 두 번째는 금융 소비자들의 이익이다. 모바일 banking, 메시지를 이용한 결제 등 다양한 금융서비스가 등장함으로써 사용자들은 금융 생활이 훨씬 편리해진다. 또한 경쟁이 치열해지면서 수수료가 낮아지거나 정교한 신용평가 모델을 통해 금융 혜택을 보지 못하던 계층에게도 기회가 주어질 수 있다. 세 번째는 일자리 창출이다. 많은 자금이 운영되고 수익성이 좋은 금융산업에 첨단 IT 기술이 결합됨으로 인해 새로운 일자리가 생겨난다. 기존 금융기업은 IT 기술자들을 필요로 하고, 새롭게 핀테크에 진출한 기업들도 새로운 비즈니스를 위해 전문가들을 필요로 한다.

핀테크 비즈니스를 하는 데 있어 클라우드는 다양한 장점을 제공한다. 기본적으로 클라우드가 가진 장점인 민첩성과 확장성은 혁신적인 아이디어를 빠르게 구현하고 안정적으로 운영하는 데 큰 도움이 된다. 또한 클라우드의 개방성은 전통적인 금융사의 시스템을 외부 핀테크 기술과 결합하는 데 유용하다. 클라우드가 제공하는 API 서비스 기능은 핀테크 기업들 간에 금융시스템들을 연동할 때 필요한 API 구현을 돕는다. 특히 오픈뱅킹의 핵심 기술은 바로 이 API 서비스이며, 클라우드가 제공하는 API Gateway를 활용하면 기존의 폐쇄적인 코어뱅킹 서비스를 API 형태로 외부에 빠르게 제공할 수 있다. 그 외에도 클라우드가 제공하는 빅데이터 분석, 인공지능 기술 등도 혁신적인 핀테크 기술을 개발을 쉽게 해 준다.

## 2 클라우드 사용에 필요한 규정

이렇게 유용한 클라우드를 핀테크 전문 기업이나 금융사가 사용하기 위해서는 전자금융감독규정에 있는 클라우드 관련된 규정을 알아야 할 필요가 있다. 여기서는 클라우드 사용에 필요한 전자금융감독규정에 대해 알아본다.

〈그림 11-26〉 전자금융 관련 규율체계



금융위원회는 2019년 1월 1일 '개정 전자금융감독규정'을 통해 클라우드 활용 범위를 확대하면서도 그것의 이용과 관련된 관리, 감독체계를 강화하였다. 개정된 규정에서, 클라우드 서비스를 활용할 수 있는 범위를 확대하고 클라우드 서비스를 이용할 때 준수해야 할 안전성과 관련된 기준을 제시하였다. 또한 핀테크 기업이 클라우드 서비스를 이용할 때 거쳐야 할 절차를 규정하고 그에 관련된 금융감독 기관의 관리, 감독 방안을 정하였다. 주요 내용은 다음과 같다.

## 2-1 클라우드 서비스 이용 범위

기존에는 금융회사가 개인신용정보와 고유식별정보가 포함되지 않은 비중요정보를 처리할 때만 클라우드 서비스를 이용할 수 있었다. 그런데 개정된 규정에서 이와 같이 중요한 데이터를 처리할 때도 클라우드를 이용 가능하도록 범위를 확대했다. 다만, 이 경우에 신속한 장애 대응과 복구가 가능하도록 국내에 위치한 클라우드 데이터센터를 통해서만 가능하도록 범위를 제약했다.

이 부분은 기존의 규정에 비해 크게 개선된 것으로, 그 전까지는 개인정보가 담긴 데이터가 조금이라도 있으면 클라우드를 전혀 사용할 수 없어 사실상 금융사가 이용 가능한 업무가 많지 않았다. 특히 클라우드의 강점인 데이터 분석이나 AI/ML 기술은 데이터를 기반으로 하기 때문에 거의 사용할 수 없다고 봐야 했으나, 이번 개정을 통해 금융사가 클라우드의 혜택을 볼 수 있는 기회가 만들어진 것이다. 국내에 위치한 데이터센터를 써야 한다는 규정도 현재 대부분의 Public 클라우드 공급업체들이 국내에 데이터센터를 가지고 있어 큰 문제가 되지 않는다. 다만, 클라우드의 일부 서비스 중에서 특정 리전(데이터 센터)에서 제공할 수 없는 글로벌 서비스(주로 AI/ML 서비스 중에서 음성인식과 같이 API 기반으로 제공하는 것들)의 경우에는 저장을 하지 않아도 처리 과정에 개인정보가 포함될 가능성이 있어 사용이 쉽지 않아진다.

## 2-2 클라우드 이용할 때 거쳐야 할 절차

주요한 절차는 다음과 같다.

- (1) 업무 중요도 평가
- (2) CSP 건전성, 안전성 평가
- (3) 정보보호위원회 심의, 의결
- (4) 감독당국 보호

첫 번째는 금융회사가 클라우드 적용 대상 업무가 중요한 업무인지 평가하는 과정이다. 중요업무의 기준은 고유식별정보나 신용정보를 다루는 경우고, 다만 전자금융거래와 관련이 없고 고객정보와 무관한 정보, 내부 직원 정보는 이에 해당하지 않는다. 중요업무에 대해서는 금융회사가 클라우드의 관리적·물리적·기술적 보호조치 등 안전성을 자체 평가(안전성 평가)하여 안전성이 확보된 클라우드만 사용하도록 하였다. 이를 위해 금융사 내에 자체적으로 정보보호위원회를 구성하여 클라우드의 안전성 평가를 심의, 의결하고 관리 감독하도록 했다. 또한 이러한 결정 내용은 감독당국에 보고해야 한다.<sup>22)</sup>

22) 「전자금융감독규정」이 정하는 '감독당국 보고'는 중요 이용의 경우에만 적용되고, 비중요이용인 경우에는 「전자금융감독규정」이 아닌 「금융회사의 정보처리 업무 위탁에 관한 규정」에 따라 보고의무가 발생한다.

금융사 자체적으로 클라우드의 안전성을 평가하고 심의하는 부분은 개별 금융사, 특히 핀테크 스타트업이 수행하기에는 어려움이 있는 부분이다. 그래서 금융사나 핀테크 기업은 금융보안원에 의뢰하여 평가를 진행할 수 있다. 얼마 전까지는 회사별로 따로 안전성 평가를 진행해야 했기 때문에 중복평가라는 논란이 있었지만, 2021년 7월 현재는 ‘합동평가’ 형태로 운영 중으로 특정 클라우드에 대한 평가를 희망하는 금융사가 한 데 모여 함께 금융보안원을 통해 안전성 평가를 진행할 수 있게 되었고(단, 대표 금융사의 참가 필요), 이에 참여하지 않은 다른 금융사도 사후에 참여 신청을 하여 평가결과를 보다 편리하게 공유받을 수 있게 되었다.

### 2-3 클라우드 서비스 안전성 확보조치 방안 수립

금융회사는 클라우드 서비스를 이용할 때에도 물리적 망분리의 등 규정상 명시적으로 정한 일부 예외를 제외하면 기존에 적용되던 동일한 안전성 확보기준을 준수해야 한다. 금융보안원의 클라우드컴퓨팅서비스 이용 가이드는 이와 관련한 내용을 안내하고 있다. 데이터 보호장치로서 외부 통신망과의 분리 및 차단 조치, 중요 정보의 암호화 처리를 규정하고, 서비스 장애 예방 및 대응에 필요한 기준도 명시하였다. 또한 클라우드 관련 장애가 발생해도 금융서비스가 중단되지 않도록 중요 전산장비를 이중화하고 백업해야 하며, 장애 발생 시 반드시 통지하고 비상 대응조치를 취하도록 하였다.

외부 통신망과의 분리 차단은 일부 물리적 망분리 예외를 제외하면 기존 금융회사에 적용되던 망분리 규정과 동일하며, 이는 클라우드 제공자 구간에 위치한 정보처리시스템에 대해 동일 클라우드 서비스를 이용하는 외부기관의 통신망과 분리 및 차단하고 접속을 금지해야 한다는 것이다. 내부망 정보처리 시스템이 클라우드 서비스 제공자 구간에 위치한 경우에는 인터넷 등 외부통신망과도 분리 및 차단하고 접속을 금지해야 한다.

## 2-4 클라우드 사업자에 대한 관리

클라우드 이용 계약서에 손해배상, 계약해지, 재판관할 사항 등을 명확하게 명시하도록 했으며, 고객 피해의 경우 금융사가 직접 배상하고 클라우드 제공자는 연대책임을 부담해야 한다. 금융회사 손해의 경우에는 클라우드 제공자의 고의, 과실로 인한 서비스 품질 저하나 장애 발생 시 클라우드 제공자가 금융회사의 손해를 배상해야 한다. 또한 금융회사나 감독당국의 조사 및 접근권(현장방문 포함)을 계약서에 명시해야 한다.

## 2-5 금융클라우드 안전성 평가

핀테크를 포함한 금융기업이 금융업무를 클라우드를 사용하기 위해서는 핀테크 업체를 포함한 금융기업이 금융업무를 클라우드를 사용하기 위해서는 금융보안원이 제시하고 있는 금융클라우드 안전성 평가(이하 안전성 평가) 기준을 바탕으로 금융사가 평가를 수행하여 준수 여부를 판단해야 한다. 아래는 금융보안원에서 제공하고 있는 가이드라인이다.

<https://www.fsec.or.kr/common/proc/fsec/bbs/147/fileDownload/1788.do>

여기에 포함된 평가와 관련된 주요 내용을 요약하면 다음과 같다.

### 평가 대상

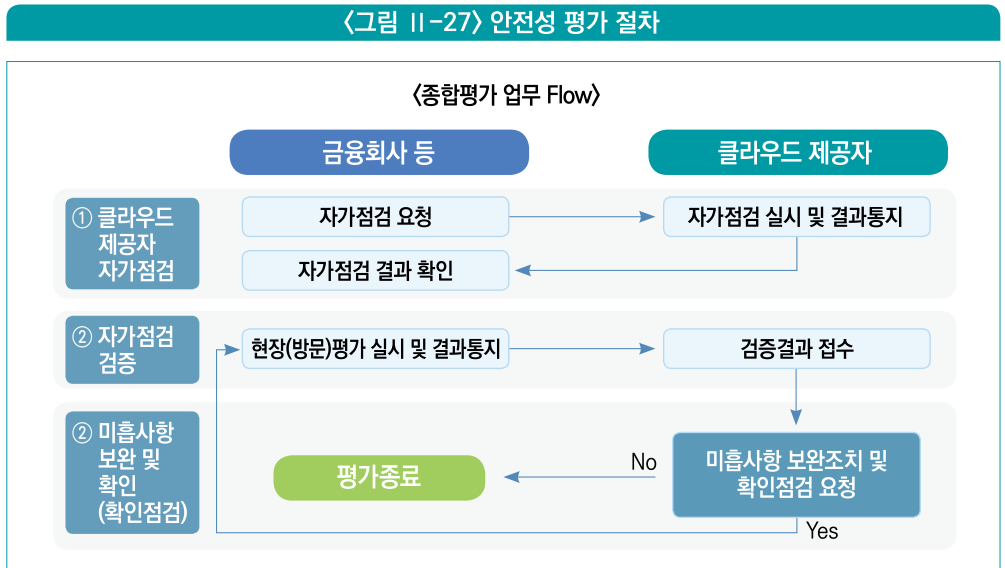
금융회사가 제3자로부터 상용으로 제공받는 모든 유형(IaaS, PaaS, SaaS)의 클라우드 서비스가 평가 대상이다. 금융회사가 자체적으로 구축하여 운영하거나, 협회나 계열사 등이 자체 구축하여 그 구성원을 대상으로 제공하는 비상용 클라우드는 평가 대상이 아니다. 이에 따르면 제3절에서 클라우드 제공자(CSP)로 언급했던 아마존 AWS, 마이크로소프트 Azure, 구글 GCP와 같은 글로벌 사업자와 네이버 NCP 등은 모두 평가 대상이다.

### 평가 절차 및 방법

평가는 크게 금융회사가 특정 클라우드의 서비스를 최초 이용 시 하는 '종합 평가'와 이용 중에 보안수준 유지 여부를 확인하기 위한 '정기 평가'가 있다. 종합 평가는 최초뿐만 아니라



같은 제공자라도 이용 서비스가 확대된 경우(IaaS → IaaS, PaaS 추가)에는 평가를 다시 진행해야 한다. 평가 절차는 다음과 같이 CSP가 자가 점검을 하고 금융회사(또는 금융보안원 대행)가 점검 내용을 검증하고 마지막으로 미흡사항에 대한 보완조치에 대한 확인 순으로 진행된다. 이때, 금융회사는 현장(방문) 평가를 통해 검증한다.



정기 평가는 CSP의 보안수준 유지 여부를 확인하기 위해 종합 평가 종료 후 매 1년마다 실시한다. 절차는 동일하며 기본적으로는 서면 검증이지만, 매 3년차 또는 금융회사가 필요하다고 판단이 될 경우 현장평가를 실시한다.

**평가 항목**

평가 항목은 CSP가 준수해야 할 일반적 보안 기준인 ‘기본 보호조치’와 금융 분야에 특화된 ‘금융 부분 추가 보호조치’로 나뉜다. 먼저 기본 보호조치는 클라우드 서비스 보안인증제(CSAP)의 평가 및 인증 기준 14개 카테고리 중 공공부분 추가 보안요구사항을 제외한 것들을 사용한다.

한편 건전성 평가는 금융회사의 전자금융감독규정 등에 따른 의무 준수를 위해 클라우드 서비스 제공자의 지원 및 협조가 필수적인 사항 9가지에 대해 점검한다.

〈표 11-2〉 CSAP의 평가인증 통제항목

통제분야	통제항목수	
	IaaS	SaaS
1. 정보보호 정책 및 조직	5개	5개
2. 인적보안	12개	5개
3. 자산관리	10개	3개
4. 서비스 공급망 관리	4개	3개
5. 침해사고 관리	7개	7개
6. 서비스 연속성 관리	7개	6개
7. 준거성	4개	3개
8. 물리적 보안	12개	-
9. 가상화 보안	10개	6개
10. 접근통제	10개	10개
11. 네트워크 보안	6개	5개
12. 데이터 보호 및 암호화	10개	8개
13. 시스템 개발 및 도입 보안	12개	10개
<b>합계</b>	<b>109개</b>	<b>71개</b>

출처: 한국인터넷진흥원(KISA)

## 1 클라우드의 주요 기술

클라우드는 다양한 기반기술들을 이용하여 구현되어 있다. 컴퓨팅 자원을 소프트웨어로 제어하는 가상 머신, 데이터를 저장하는 스토리지를 구현하는 분산 스토리지 및 스토리지 가상화 기술, 그리고 데이터를 원격으로 주고 받을 수 있는 네트워크 기술 등이 그것이다. 다음 <그림 11-28>과 같이 클라우드 기술은 가장 하부에 있는 물리적인 자원들 즉, 컴퓨터, 스토리지, 네트워크 등의 물리적인 자원들을 가상화(virtualization) 기술을 통해 소프트웨어화(물리적인 자원들을 직접 물리적인 동작을 통해 조작하지 않고 소프트웨어로 다루는 것을 말한다. 완전히 동일하지는 않지만 예전 자동차에서 기어 조작이 직접적인 힘을 스틱을 통해 기계에 전달하여 기어를 변경하는 것이었다면, 최근에는 버튼만 가지고 물리적인 조작 없이 전기 신호로 제어하는 것과 유사하다고 할 수 있다)하고 그것을 클라우드 서비스 계층을 통해 사용자들이 쉽게 자원들을 사용할 수 있게 한다.

〈그림 11-28〉 클라우드 기술 스택

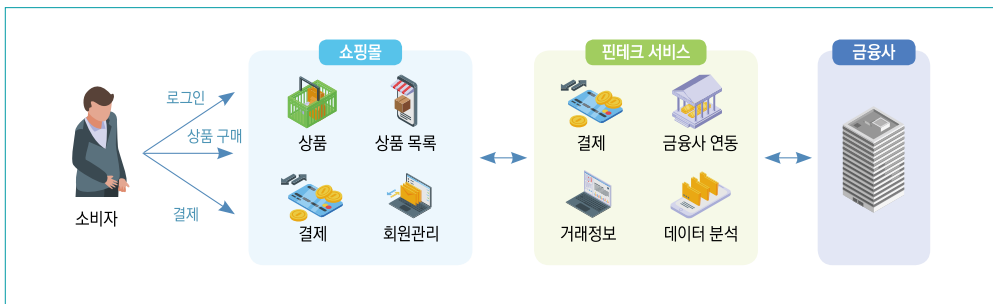


물리적인 자원을 직접 다루어야 한다면 컴퓨터나 스토리지, 네트워크 장비들을 직접 물리적으로 설치하고 서로 연결해야 한다. 이것은 물리적인 행위가 필요하기 때문에 많은 시간이 걸리고 실제 조작 과정이 번거롭다. 그러나 가상화를 통해 소프트웨어화 하게 되면, 직접 물리적으로 동작할 필요 없이 단지 실행 버튼이나 API(Application Programming Interface) 호출을 통해서 편리하게 제어할 수 있다. 그러면, 직접 데이터센터에 가서 컴퓨터 전원을 만지지 않고도 가상 머신(가상화된 컴퓨터)을 켜거나 끌 수 있다. 네트워크 회선을 직접 연결하지 않아도 서로 다른 가상 머신 간에 네트워크 통신을 연결할 수 있다. 이것이 클라우드의 주요 기술인 가상화 기술의 장점이다.

클라우드에는 여기에 그치지 않고 소프트웨어화 된 가상자원들을 UI 화면을 통해 쉽게 제어할 수 있도록 클라우드 서비스들을 제공한다. 클라우드에서 사용할 수 있는 클라우드 자원들을 정의하고 목록화(카탈로그)하여 사용자가 쉽게 찾을 수 있게 한다. 계정 관리를 통해 여러 사용자가 동시에 접근해도 서로 완벽하게 분리하여 자원이나 데이터가 섞이는 것을 방지한다. 이러한 것들은 서비스 포털이 제공하는 UI를 통해 사용할 수 있다.

이 절에서는 이러한 각각의 클라우드 기술에 대해 알아보고, 핀테크 서비스를 구현한다면 어떤 클라우드 서비스들을 사용하게 될지 가상의 업무 시나리오에 맞춰 클라우드 서비스를 구축해 보도록 하겠다. 가상의 업무 시나리오는 다음과 같다. 소비자가 쇼핑물을 통해 상품을 구매하고 결제를 핀테크 서비스에 요청한다. 핀테크 서비스는 금융사들과 연계하여 카드결제나 송금 등의 금융 업무를 진행한다. 쇼핑물은 결제가 완료된 상품을 소비자에게 전달한다. 여기서 쇼핑물은 핀테크 서비스의 활용 사례로 넣은 것이므로 실제 클라우드 서비스 구축 범위는 중간에 있는 핀테크 서비스로 한정하겠다.

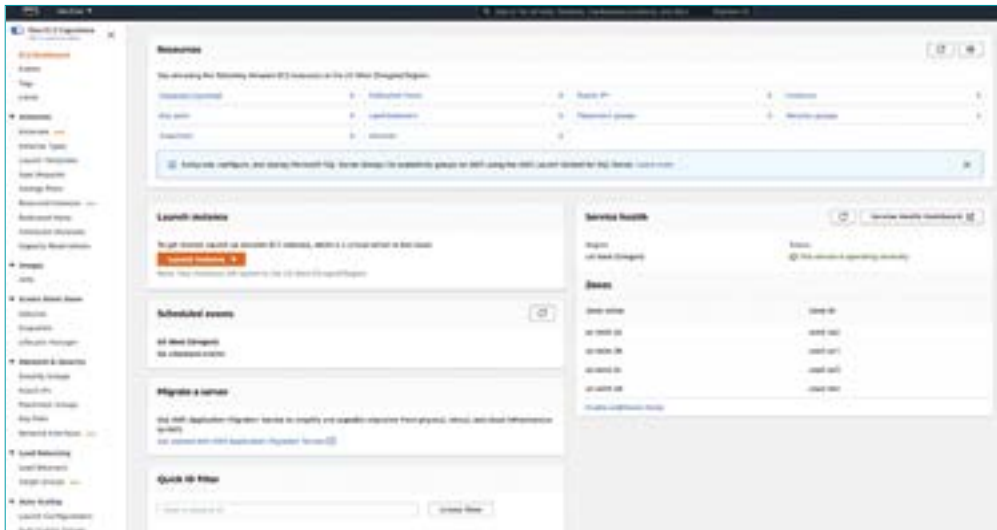
〈그림 11-29〉 가상의 핀테크 서비스



## 2 컴퓨팅 혹은 가상 머신(IaaS)

컴퓨팅 혹은 가상 머신은 소프트웨어를 동작할 수 있는 환경을 의미한다. 집에 있는 노트북이나 데스크탑은 대표적인 컴퓨팅 자원이라 할 수 있다. 노트북은 특정 CPU를 사용하고 보통 여러 개의 코어이며 일정 크기의 메모리를 가지고 있다. 클라우드는 가상화 기반으로 컴퓨팅 환경을 다양하게 제공한다. 각 클라우드별로 개념은 동일하지만 서로 다른 이름을 가지고 있는데, 아마존 AWS에서는 EC2(Elastic Compute Cloud), MS 애저에서는 Azure Compute, 구글 GCP에서는 GCE(Google Compute Engine)이라고 부른다.

〈그림 11-30〉 아마존 AWS의 컴퓨팅 서비스인 EC2



출처: AWS

클라우드에서 제공하는 컴퓨팅 자원은 가상화를 통해 매우 다양한 스펙의 가상 머신을 제공한다. 구글 GCP를 예를 들어보면 먼저 가장 무난하게 사용할 수 있는 N1, N2 타입의 가상 머신을 제공한다. 이러한 타입의 가상 머신은 빠른 CPU에 충분한 메모리를 제공하여 대부분의 업무에 적합하다. 하지만, 만약 스타트업에서 적은 비용으로 서비스를 구축하고 싶다면 가격대 성능비가 좋은 E2 타입의 가상 머신을 사용할 수 있다. E2 타입의 가상 머신을 사용하면 복잡한 트랜잭션과 같은 무거운 워크로드보다는 웹서비스, 가상 데스크탑과 같은 상대적으로 가벼운 업무를 돌리는 데 적합하다.

만약 데이터베이스나 ERP 등과 같은 무거운 워크로드이면서도 메모리를 많이 사용하는 업무를 돌리는 경우라면 M1, M2와 같은 메모리 최적화 타입을 사용하는 것이 좋다. 이러한 타입의 가상 머신에는 범용 머신에 비해서 높은 용량의 메모리를 탑재한다. 예를 들어 범용 N1, N2 가상 머신에서는 128GB에서 1TB 정도의 메모리를 최대로 제공하는데 메모리 최적화 M1, M2 가상 머신에서는 4TB에서 12TB 용량(2021년 7월 현재)의 메모리까지도 제공한다. 참고로 SAP와 같은 ERP를 돌릴 때는 수 TB부터 많으면 수십 TB까지의 메모리를 필요로 하기도 한다.

〈표 II-3〉 구글 GCP의 Compute 종류

범용		워크로드 최적화		
비용 최적화	균형	메모리 최적화	컴퓨팅 최적화	가속기 최적화
E2	N2, N2D, N1	M2, M1	C2	A2
저렴한 비용으로 일상적인 컴퓨팅	다양한 VM 형태에서 균형 있는 가성비 제공	메모리 사용량이 매우 큰 워크로드	컴퓨팅 집약적 워크로드를 위한 초고성능	고성능 컴퓨팅 워크로드에 최적화
<ul style="list-style-type: none"> <li>• 웹 서비스</li> <li>• 앱 서비스</li> <li>• 백오피스 앱</li> <li>• 중소 규모 데이터 베이스</li> <li>• 마이크로서비스</li> <li>• 가상 데스크톱</li> <li>• 개발 환경</li> </ul>	<ul style="list-style-type: none"> <li>• 웹 서비스</li> <li>• 앱 서비스</li> <li>• 백오피스 앱</li> <li>• 중대형 데이터베이스</li> <li>• 캐시</li> <li>• 미디어/스트리밍</li> </ul>	<ul style="list-style-type: none"> <li>• SAP HANA와 같은 중대형 인메모리 데이터베이스</li> <li>• 인메모리 데이터베이스 및 인메모리 분석</li> <li>• Microsoft SQL server 및 유사 데이터베이스</li> </ul>	<ul style="list-style-type: none"> <li>• 컴퓨팅 제약 워크로드</li> <li>• 고성능 웹 서비스</li> <li>• 게임(AAA 게임 서버)</li> <li>• 광고계제</li> <li>• 고성능 컴퓨팅(HPC)</li> <li>• 미디어 트랜스코딩</li> <li>• AI/ML</li> </ul>	<ul style="list-style-type: none"> <li>• CUDA 사용 설정 ML 학습 및 추론</li> <li>• HPC: 고성능 컴퓨팅</li> <li>• 동시에 로드되는 대규모 계산</li> </ul>

만약 게임이나 고성능 연산이 필요한 업무라면 컴퓨팅 최적화 타입의 가상 머신인 C2를 사용할 수 있다. 이러한 가상 머신에는 클럭이 높은 고성능 CPU가 들어 있어 같은 숫자의 코어라 하더라도 더 높은 연산 처리 능력을 낼 수 있다. 참고로 게임사에서는 수십 ~ 수백 대의 하드웨어를 이용하여 게임 서비스를 하는데, 고성능 머신을 사용하여 작은 숫자의 하드웨어를 이용하는 것이 관리나 비용 측면에서 유리하기 때문에 컴퓨팅 최적화 머신을 많이 사용한다.

AI/ML 워크로드를 돌리기 위해서는 CPU보다는 GPU가 유리하다. 그렇다고 GPU 단독으로 돌아갈 수는 없으므로 CPU가 있는 일반 가상 머신에 GPU를 연결하는데, 가속기 최적화 타입의 머신에서는 최신 A100 GPU를 사용할 수 있도록 제공한다. 구글 GCP에서는 범용 타입이나 컴퓨팅, 메모리 최적화 머신에 모두 GPU(K80, P100, V100 등)를 연결하는 기능도 제공하지만, 가속기 최적화 머신을 이용하여 최고 사양의 A100을 사용할 수도 있는 것이다.

범용 또는 컴퓨팅, 메모리 최적화 유형 중에서 선택을 하게 되면 다음은 어느 정도의 CPU와 메모리를 가진 가상 머신을 사용할지 선택하게 되는데, 통상 다음과 같이 CPU 개수와 메모리 크기가 어느 정도 정해진 형태로 서비스를 제공한다. 즉, 4 코어에서는 16GB, 8 코어에서는 32GB 등이 그것이다. 그러나 구글 GCP 같은 경우에는 사용자가 코어수와 메모리 크기의 조합을 다양하게 변경할 수 있는 커스텀 타입도 제공하고 있다. 어떤 정도의 코어나 메모리가 필요한지는 정해진 규칙은 없다. 자동차를 살 때 몇 마력이 정답인지, 에어컨 몇 평형이 정답인지 정해진 것은 없듯이 클라우드 컴퓨팅 서비스에서도 정답은 없다. 보통은 서비스 구현 과정에 성능 테스트를 진행해서 적정 크기를 산정한다. 하지만 클라우드의 장점은 이러한 머신의 크기를 한 번에 결정할 필요 없이 필요할 때 즉시 마우스 클릭만으로 증설하거나 부하에 따라 자동으로 용량을 조절하는 오토 스케일링 기술을 사용할 수 있으므로 자동차나 에어컨 선택보다는 더 용이하다고 볼 수 있다.

〈표 II-4〉 구글 GCP의 범용 가상 머신 목록

N2 표준    N2 높은 메모리    N2 높은 CPU

N2 표준 머신에는 vCPU당 4GB의 시스템 메모리가 있습니다.

머신 유형	vCPU <sup>1</sup>	메모리 (GB)	최대 영구 디스크 수(PD) <sup>2</sup>	최대 총 PD 크기(TB)	로컬 SSD	기본 이그레스 대역폭(Gbps) <sup>3</sup>	등급 1 이그레스 대역폭(Gbps) <sup>4</sup>
n2-standard-2	2	8	128	257	있음	10	해당 사항 없음
n2-standard-4	4	16	128	257	있음	10	해당 사항 없음
n2-standard-8	8	32	128	257	있음	16	해당 사항 없음
n2-standard-16	16	64	128	257	있음	32	해당 사항 없음
n2-standard-32	32	128	128	257	있음	32	50
n2-standard-48	48	192	128	257	있음	32	50
n2-standard-64	64	256	128	257	있음	32	75
n2-standard-80	80	320	128	257	있음	32	100

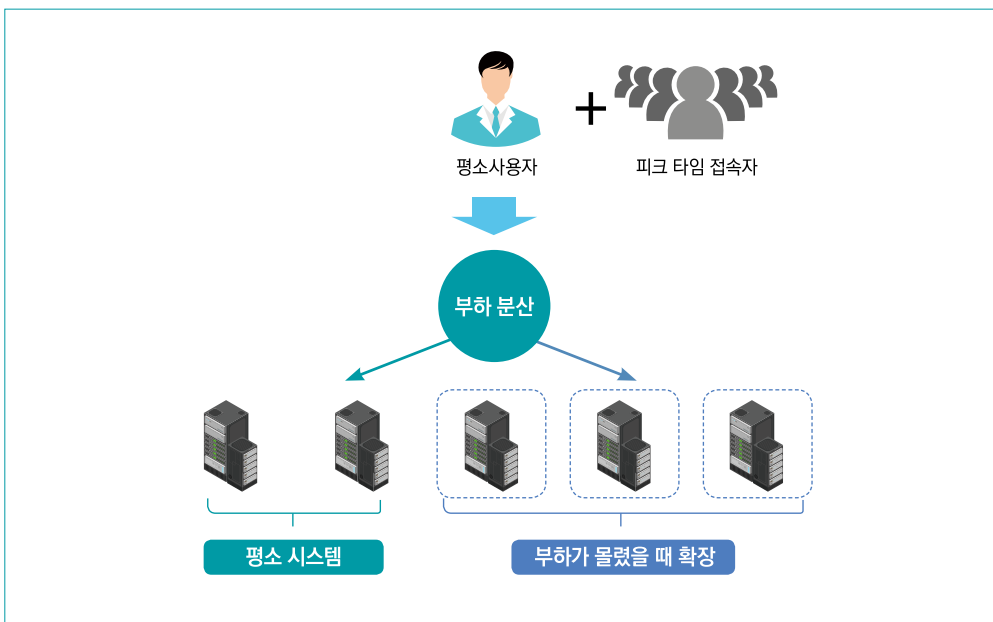
<sup>1</sup>vCPU는 사용 가능한 CPU 플랫폼 중 하나에서 단일 하드웨어 하이퍼 스레드로 구현됩니다.  
<sup>2</sup>영구 디스크 사용 요금은 머신 유형 가격과 별도로 청구됩니다.  
<sup>3</sup>최대 이그레스 대역폭은 지정된 수를 초과할 수 없습니다. 실제 이그레스 대역폭은 대상 IP 주소와 기타 요인에 따라 다릅니다. 네트워크 대역폭을 참조하세요.  
<sup>4</sup>32Gbps Skylake 이상 CPU 플랫폼. 다른 모든 플랫폼에서 10Gbps

출처: 구글 클라우드 플랫폼 홈페이지



제4절의 쇼핑몰 사례에서 오토 스케일링 기능에 대해 설명했는데, 이것은 컴퓨팅 자원의 용량을 피크 타임에 맞추어 최대로 준비하는 것이 아니라 평소에는 최소한의 자원을 유지하다가 급격히 사용자가 몰리면 그에 맞추어 컴퓨팅 자원의 숫자를 늘리는 것이다. 여기서 주의할 점은 머신의 CPU나 메모리를 증설하는 방식(scale-up)이 아니라 동일한 사용인 머신의 숫자를 늘리는 방식(scale-out)이라는 점에 주의하고 소프트웨어 구현 시 이러한 부분을 감안하여 약간의 아키텍처 고려를 해야 한다. 클라우드상에서 컴퓨팅 자원을 이용하여 서비스를 구축하는 경우에는 거의 대부분 오토 스케일링을 이용한다는 점은 명심해 두자.

〈그림 II-31〉 auto-scaling 기능

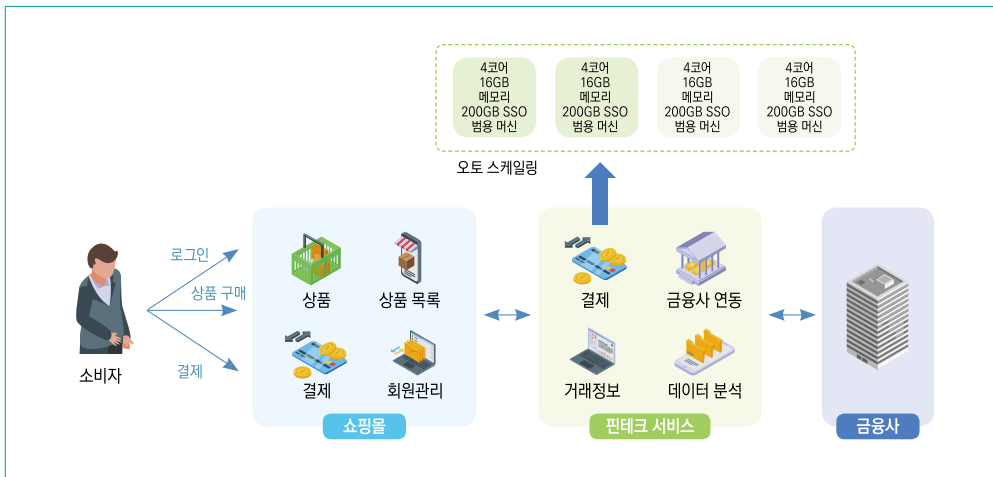


컴퓨팅 자원에서 CPU, 메모리와 함께 같이 필요한 자원은 디스크이다. 다음에 클라우드의 스토리지 서비스가 있지만, 컴퓨팅 자원에 사용되는 디스크는 그것과는 다른 것이므로 주의하자. 쉽게 생각하면 컴퓨팅 자원에 연결된 디스크는 노트북이나 데스크탑에 들어 있는 하드디스크 또는 SSD라고 생각하면 되고, 스토리지 서비스는 더 큰 자료를 저장하고 위해

구글 드라이브(<http://drive.google.com>)나 드랍박스(<http://www.dropbox.com>)를 사용하는 것과 유사하다. 디스크는 보통 하드디스크와 유사한 일반 디스크와 반도체 메모리를 이용한 SSD로 나눌 수 있다. 일반적으로 500GB 내외의 작은 용량으로 빠른 서비스에 필요한 데이터는 SSD를 사용하고, 1TB 이상의 대용량 데이터를 굳이 빠르게 액세스할 필요가 없다면 일반 디스크를 권장한다. 다만, 데이터베이스처럼 고성능 데이터 처리가 필요하면 전체를 SSD로 사용하는 것을 추천한다. 참고로 클라우드에서 제공하는 스토리지들은 동적으로 용량을 확장할 수 있다. 우리가 노트북을 구입할 때 내장된 디스크의 용량이 모자라면 다시 구매해서 교체하지 않는 이상 용량을 늘릴 수 없는 반면, 클라우드의 디스크는 소프트웨어적인 가상화를 통해서 제공하기 때문에 단지 클라우드 UI에서 숫자로 된 용량을 바꾸는 것 만으로도 증설이 가능하다.

그럼 가상의 핀테크 서비스에서 적절한 컴퓨팅 자원을 선정해 보자. 먼저 결제 서비스의 경우 쇼핑몰에서 오는 결제 요청을 받아서 처리해야 하므로 웹서비스와 결제 소프트웨어를 운영해야 한다. 여기에 컴퓨팅 자원을 사용하고 필요한 소프트웨어를 설치 또는 개발하기로 한다. 일반적인 워크로드이므로 4코어에 16GB 범용 머신 두 대를 최소로 서비스하고 필요시 부하에 따라 머신의 개수를 늘릴 수 있도록 오토 스케일링을 적용한다. 통상 오토 스케일링으로 증가할 수 있는 머신의 개수에 제약을 둘 필요가 없으므로 여기서도 필요한 만큼 늘어날 수 있도록 최대 머신 개수 제한은 두지 않기로 한다. 디스크는 결제 서비스에서 특별히 많은 데이터를 저장할 필요가 없으므로 200GB SSD를 사용한다. 만약 미디어 파일 등을 대량으로 저장해야 한다면 1TB 이상의 일반 디스크를 사용하는 것이 좋겠지만 지금 여기서의 핀테크 시나리오에서는 200GB 정도면 충분할 듯 하고, 대신 SSD를 써서 서비스의 속도를 올리도록 한다.

〈그림 11-32〉 가상의 핀테크 서비스



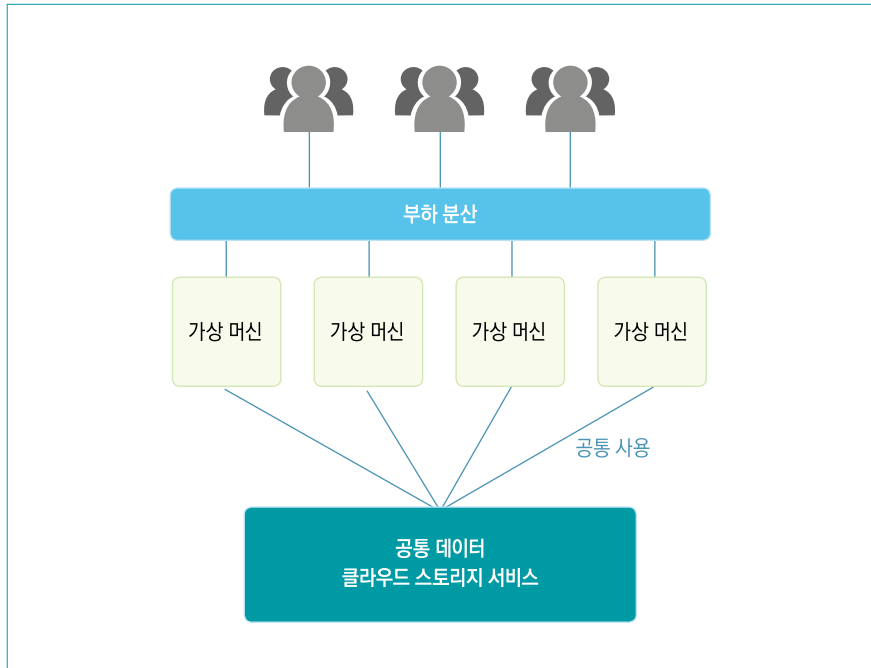
### 3 스토리지

클라우드 스토리지 서비스는 디스크처럼 특정한 가상 머신에 연결되거나 내장된 스토리지가 아니다. 대신 여러 가상 머신이 공통적으로 사용할 수 있는 저장 공간이다. 마치 회사 내에 있는 공유 폴더나 구글 드라이브 같은 클라우드 서비스라고 보는 것이 맞다. 실제로도 클라우드 스토리지의 역할은 공유 폴더처럼 여러 애플리케이션이나 서비스들이 공통의 자료를 저장하거나 대량의 데이터를 저장할 때 사용하는 서비스이다. 클라우드 스토리지를 사용하는 대표적인 경우는 미디어 서비스에서 영상 파일들을 저장할 때이다. 또는 빅데이터 분석 업무에서 대량의 데이터를 공유 스토리지에 저장해 놓고 분석가들이 원할 때 가져다 쓰는 경우에도 유용하다. 이 외에도 프로그램이나 소프트웨어에서 발생하는 로그를 저장하는 경우에도 종종 사용한다.

클라우드 스토리지의 장점은 크게 두 가지로 볼 수 있다. 첫 번째는 여러 가상 머신이나 서비스들이 공동으로 사용할 수 있는 서비스이기 때문에, 공통으로 필요한 데이터들을 저장에 놓고 사용할 수 있다. 두 번째로 클라우드 스토리지는 거의 무한대에 가까운 저장

용량을 가진다는 점이다. 사실 클라우드 스토리지 서비스를 사용하는 가장 큰 이유 중의 하나가 바로 이 무한의 용량을 가진다는 점이다. 그것은 컴퓨팅 자원에 연결되는 디스크가 일정 용량의 저장 공간을 가지는 데 반해 클라우드 스토리지는 클라우드 사업자의 용량 한계에 걸리지 않는다면 원하는 만큼 자유롭게 사용할 수 있다. 특히 비용도 디스크가 데이터를 저장하지 않아도 확보한 용량만큼 비용을 지불하는 반면, 클라우드 스토리지는 철저히 종량제로 저장된 데이터의 양에 비례해서 과금된다.

〈그림 11-33〉 클라우드 스토리지 사용 사례



컴퓨팅 자원에 연결된 디스크는 운영체제에서 바로 사용 가능한(윈도우에서는 탐색기에서 파일을 사용할 수 있음) 반면, 클라우드 스토리지는 특별한 명령이나 API를 사용해야 한다. 구글 드라이브나 드랍박스가 별도의 웹서비스나 설치 프로그램을 사용하는 것과 유사하다. 다만, 구글 드라이브 등의 서비스는 사용자의 편의를 위해 동기화 기능을 통해 사용을

쉽게 해주고 있지만, 클라우드 서비스는 그런 동기화 서비스는 제공하지 않는다. 일반적으로 클라우드 기반의 프로그램을 작성하는 개발자들은 이러한 클라우드 클라우드 업체가 제공하는 명령이나 API를 통해 스토리지 서비스를 사용하는 방법을 잘 알고 있으므로 큰 문제는 없다. 클라우드 스토리지가 무한대에 가까운 저장소를 제공하는 면에서 장점은 있지만, 반대로 그러다 보니 그것을 사용하는 데는 약간의 작업이 필요하다는 것을 알 수 있다.

클라우드 스토리지는 대용량(테라바이트를 넘어 페타바이트 이상의 용량을 저장하기도 함)의 데이터를 저장하다 보니, 아무리 클라우드가 저렴하다 하더라도 비싼 비용을 내야 하기도 한다. 반면, 사용자 입장에서는 비용 걱정 없이 데이터를 마음껏 저장하고 싶을 것이다. 클라우드 업체들은 이러한 사용자의 수요를 맞추기 위해 스토리지 티어(tier)라는 것을 제공한다. 예를 들어 구글 GCP는 네 가지 클래스(티어와 유사한 개념)의 스토리지 서비스를 제공한다. standard, nearline, coldline, archive가 그것이다. 타 클라우드에서도 매우 유사한 티어 개념을 제공하고 있다.

〈표 II-5〉 구글 GCP의 스토리지 티어와 가격

스토리지 클래스	API 및 gsutil에서 사용되는 이름	최소 스토리지 기간	일반적인 유틸리티
표준 스토리지	STANDARD	없음	<ul style="list-style-type: none"> <li>업타임 리전 및 아웃 리전에서 99.99% 이상</li> <li>리전에서 99.99%</li> </ul>
Nearline Storage	NEARLINE	30일	<ul style="list-style-type: none"> <li>업타임 리전 및 아웃 리전에서 99.95% 이상</li> <li>리전에서 99.9%</li> </ul>
Coldline Storage	COLDLINE	90일	<ul style="list-style-type: none"> <li>업타임 리전 및 아웃 리전에서 99.95% 이상</li> <li>리전에서 99.9%</li> </ul>
Archive Storage	ARCHIVE	365일	<ul style="list-style-type: none"> <li>업타임 리전 및 아웃 리전에서 99.95% 이상</li> <li>리전에서 99.9%</li> </ul>

클라우드 스토리지	NEARLINE STORAGE	COLDLINE STORAGE	ARCHIVE STORAGE
10GB당 월별 \$0.02부터	10GB당 월별 \$0.01부터	10GB당 월별 \$0.004부터	10GB당 월별 \$0.0012부터

출처: 구글 클라우드 플랫폼 홈페이지

구글 GCP의 standard는 표준 스토리지 티어로서 처음에 객체를 만들었을 때 디폴트로 사용되는 것이다. 데이터 저장과 삭제가 자유롭다. 반면 위의 <표 II-5>에서 알 수 있듯이 가격은 다른 티어에 비해서 가장 비싸다. 아무런 제약 없이 자유롭게 사용할 수 있기 때문에 비용이라 생각하면 된다.

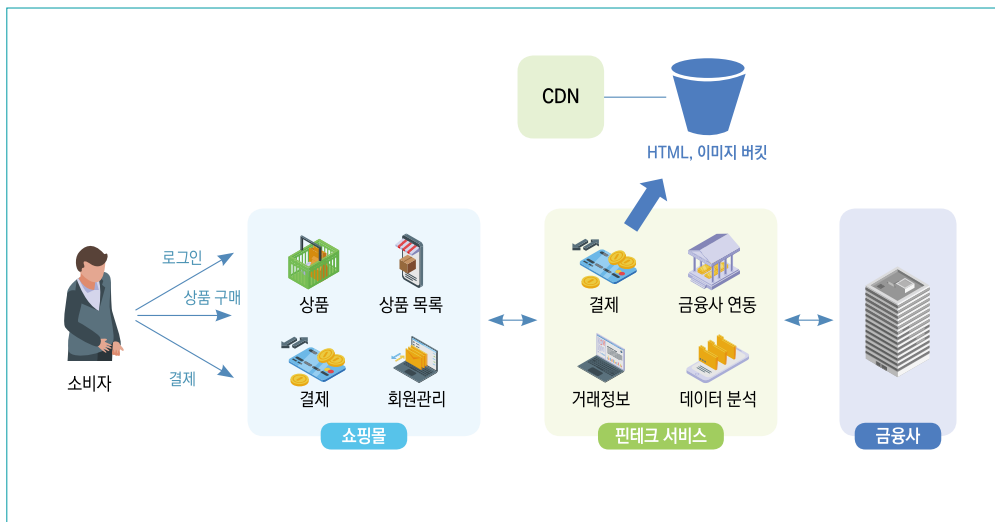
구글 GCP에서 nearline의 가격은 standard의 절반이다. 같은 가격으로 두 배의 데이터를 저장할 수 있는 것이다. 반면 제약사항이 하나가 붙는데, 그것은 데이터를 한 번 저장하면 바로 삭제를 하더라도 standard의 절반 가격으로 30일 비용이 청구된다는 점이다. 쉽게 설명하면 이런 것이다. 1000GB의 데이터를 standard와 nearline에 각각 30일을 저장한다면 각각 약 20\$와 10\$의 비용이 청구되어 nearline이 훨씬 유리하다. 다만, 데이터를 3일 지나서 삭제했다면 standard는 1/10인 2\$만 청구되는 반면 nearline에서는 그대로 30일 가격인 10\$가 청구된다. 결국 데이터를 자주 삭제하는 경우에는 standard가 유리하고 한 번 저장하면 30일 이상 삭제할 일이 없다면 nearline이 유리한 것이다. coldline에서는 다시 nearline의 40% 가격이지만 90일이라는 제약이 있고, archive는 standard의 1/20 수준이지만 1년이라는 제약이 붙는다. 처음에는 다소 혼란스러울 수 있지만 통상 90% 이상의 데이터는 자주 삭제할 일이 없으므로 nearline이나 coldline이 유리하고, 그중에서도 1년 이상 장기저장이나 백업 목적이라면 archive가 적절한 선택이라 할 수 있다. 구글 GCP를 포함한 많은 클라우드 서비스에서는 저장된 데이터가 지정된 날짜가 지나면 자동적으로 스토리지 티어를 변경(standard ⇒ nearline 또는 nearline ⇒ coldline)하는 라이프사이클 관리 기능을 제공하므로 일일이 데이터 별로 티어를 지정하지 않아도 된다.

노트북에서 여러 개의 디스크를 사용하면 통상 C 드라이브, D 드라이브 등과 같이 구분해서 사용한다. 그러나 클라우드 스토리지는 많으면 페타바이트 이상의 데이터도 저장한다고 했다. 그리고 여러 머신과 서비스 그리고 여러 사용자들이 접근을 하기 때문에 C, D 등의 드라이브만으로는 관리하기가 어렵다. 클라우드 스토리지에서는 일반적으로 버킷이라는 단위로 스토리지를 관리한다. 버킷이라는 것을 만들면 고유한 이름이 하나 필요하다. 그리고 각 버킷마다 고유한 이름이 있기 때문에 다른 서비스, 다른 사용자, 다른

기업에도 쉽게 공유를 할 수 있다. 권한부여도 버킷별로 다르게 줄 수 있어 관리하기가 편하다. 또한 버킷을 용도별로 만들어 관리할 수도 있다.

가상의 핀테크 서비스에서는 클라우드 스토리지 서비스를 어디에 사용하는 게 좋을까? 웹페이지를 통해 사용자 UI를 제공한다면 웹서버 구축에 필요한 HTML이나 이미지 등을 클라우드 스토리지에 저장하면 좋다. 보통 HTML이나 이미지들은 용량은 크지 않지만 여러 컴퓨팅 머신에서 동시에 필요하므로 클라우드 스토리지에 넣어두면 변경이 생겼을 때 여러 군데를 고치지 않을 수 있으므로 편리하다. 그리고 클라우드에서는 보통 CDN 서비스(제4절 넷플릭스 사례 참고)를 통해 스토리지에 들어 있는 HTML이나 이미지들을 바로 웹으로 서비스할 수 있다. 향후 데이터 분석을 하게 되면 분석용 빅데이터를 저장하는 용도로 사용할 수도 있을 듯하다.

〈그림 11-34〉 가상의 핀테크 서비스



참고로, 클라우드 스토리지 서비스 중에서 지금까지 말한 무한대의 스토리지 말고 정말로 공유 폴더 개념 형태로 사용할 수 있는 서비스도 있다. 아마존 AWS의 경우 EFS(Elastic File System), 구글 GCP의 경우에는 Cloud Filestore가 그것이다. 이러한 서비스들은

정말로 공유 폴더 목적으로 사용할 수 있는 스토리지를 제공한다. 앞에서 말한 무한대 확장이 가능한 클라우드 스토리지 서비스에 비해 이러한 스토리지는 보통 NAS(Network Access Storage) 서비스라고 부르는데, 정해진 크기의 스토리지를 만들어 각 컴퓨팅 머신에 마운트(mount, 윈도우의 경우 네트워크 드라이브)하여 마치 로컬 디스크인 것처럼 사용하게 해 준다. 클라우드 스토리지와 NAS 서비스의 가장 큰 차이는 용량이다. NAS 서비스의 경우 1TB~10TB 정도의 용량을 가진 데이터를 공유한다고 할 때 유리하다. 별도의 명령이나 API 구현이 필요한 클라우드 스토리지와 달리 운영체제에서 바로 사용할 수 있기 때문에 쉽게 사용할 수 있는 점도 장점이다.

## 4 데이터베이스

데이터베이스는 클라우드 기반의 업무 서비스에서 발생하는 데이터를 저장하기 위한 기능이다. 앞의 클라우드 스토리지도 데이터를 저장하기 위한 서비스라고 얘기했는데 무슨 차이가 있을 것일까? 쉽게 생각하면 클라우드 스토리지는 파일로 되어 있는 데이터를 저장하는 곳이고, 데이터베이스는 회원 정보, 거래 내역 등의 데이터를 저장하는 곳이라 보면 된다. 그래서 데이터베이스에 저장되는 데이터는 주로 정형(structured) 데이터라고 생각하면 된다. 아래 <표 11-6>은 정형 데이터의 일종인 회원 정보의 예시이다.

<표 11-6> 정형 데이터 예시 - 회원 정보

Name	Gender	Address	Telephone
홍길동	M	서울시 종로구	111-2222-3333
이순신	M	전라남도 여수시	444-5555-6666
신사임당	F	강원도 강릉시	777-8888-9999

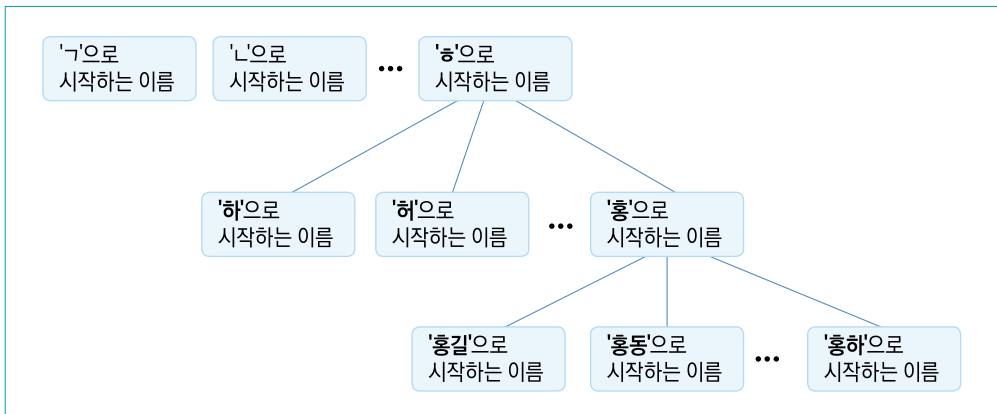
만약 회원 정보를 비정형 데이터의 일종인 텍스트(text) 파일로 만들어 클라우드 스토리지에 저장했다면 어떻게 될까? 만약 10만 명의 회원이 있다면 텍스트의 내용



중에서 '이순신' 회원을 찾아 성별, 주소 등과 같은 개인 프로필 정보를 조회하려면 텍스트 파일을 일일이 뒤져 해당 데이터를 찾아야 한다. 특히 파일을 열어서 보려면 전체 텍스트 파일을 가져와야만 한다. 회원수가 만약 100만명 또는 1000만명이 된다면 텍스트 파일을 여는 것조차 버거울 것이다. 반면, 데이터베이스는 하나의 회원 정보를 열(row)로 나누어 저장하고, 프로필의 각 필드들도 행(column)으로 나누어 저장한다. 마치 바둑판이 19\*19 구조로 되어 있고 종간의 어느 지점을 찾으려면 (3, 4)나 (16, 4)와 같이 좌표만 입력하듯이 데이터베이스에 들어 있는 데이터들은 전체를 열지 않아도 해당 데이터의 위치로 쉽게 찾아갈 수 있다. 100만명 회원정보 데이터를 각각 파일로 저장해서 찾는 것과 유사한 효과를 낸다. 그래서 정형 데이터를 저장할 때는 거의 대부분 데이터베이스를 이용한다.

여기에 추가로 데이터베이스는 인덱스라는 기능을 가지고 있어 특정 데이터를 빠르게 찾을 수 있는 기능을 제공한다. 다음 <그림 11-35>는 데이터베이스의 인덱스를 표현한 그림이다. 100만명의 회원이 데이터베이스 있고 이름을 기준으로 '홍길동'을 조회해야 한다고 가정해 보자. 인덱스를 사용하면 이름들을 먼저 가나다 순서로 정렬을 한다. 그리고 나서 'ㄱ'으로 시작하는 이름이 어느 행부터 어느 행까지 기록해 둔다. 그리고 'ㄴ', 'ㄷ' 순서로 'ㅎ'까지 저장한다. 그러면 '홍길동'을 검색할 때는 'ㅎ'으로 시작하는 이름의 위치부터 검색하는 것이 훨씬 빠를 것이다. 만약 'ㅎ'으로 시작하는 이름도 많다면 다시 '하', '허', ... '홍'의 글자 순서로 된 위치를 기록해 둔다. 그러면 'ㅎ'으로 시작되는 이름이 많아도 빠르게 '홍'을 찾아갈 수 있다. 만약 '홍'으로 시작하는 이름이 많으면 다시 '홍길'로 시작하는 위치를 기록해서 더 빠르게 찾아갈 수 있다. 내부적으로는 이보다 더 복잡한 방식으로 저장하지만 검색의 효율을 엄청나게 향상시킬 수 있다. 물론 텍스트 파일에서도 이렇게 위치를 기록하는 방식으로 빨리 찾을 수도 있겠지만 회원 데이터가 계속 추가되고 삭제된다면 매번 기록된 위치를 수정하는 수고가 뒤따를 것이다. 반면 데이터베이스는 매번 데이터를 넣을 때 자동으로 이러한 인덱스 데이터를 조정해서 번거로움을 피할 수 있다. 즉, 데이터베이스는 이런 식으로 정형화된 데이터의 저장과 검색에 효율된 데이터 저장소라고 생각하면 된다. 클라우드 스토리지는 그림, 동영상 파일처럼 길이가 일정하지 않고 인덱스를 만들기 어려운 데이터를 저장하는 효율적이지만 정형화된 데이터의 저장에는 그다지 효율이지 않다.

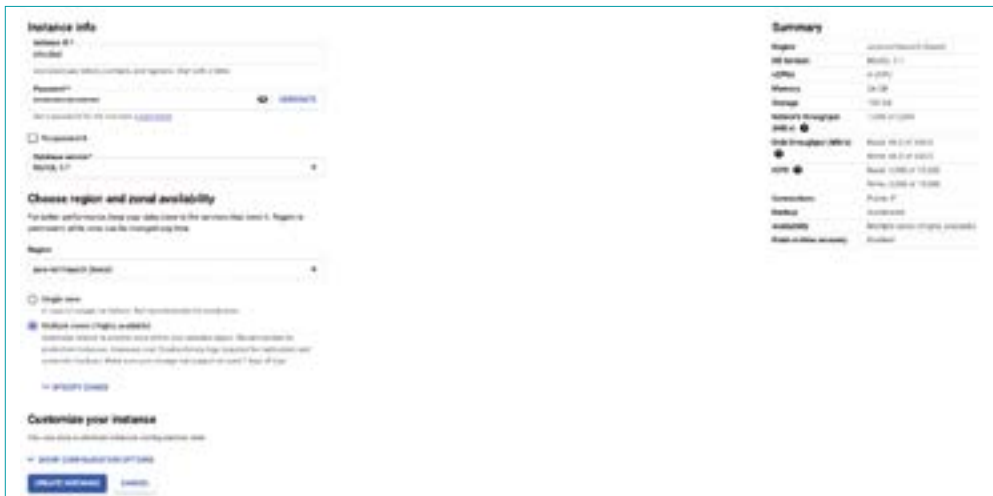
〈그림 11-35〉 데이터베이스 인덱스의 개념



데이터베이스는 이러한 정형 데이터를 100만 개, 1000만 개 혹은 그 이상 저장해도 거의 속도 저하 없이 데이터를 검색할 수 있는 장점이 있다. 특히 파일을 저장하는 클라우드 스토리지보다 훨씬 고효율의 시스템을 사용하기 때문에 매우 우수한 성능을 보인다. 반면 이러한 고성능 시스템과 복잡한 인덱스 기능을 제공하고 있기 때문에 가격 면에서는 클라우드 스토리지보다 비싸다. 그럼에도 정형화된 데이터의 저장에는 데이터베이스가 훨씬 효율적이다.

클라우드에서 이러한 데이터베이스를 사용하는 방법이나 제공하는 데이터베이스 서비스 종류는 여러 가지가 있다. 원하는 데이터베이스를 직접 컴퓨팅 VM에 설치하는 방식으로 IaaS 형태로 사용할 수도 있고 클라우드가 설치나 운영까지 알아서 해주는 PaaS 방식으로 사용할 수 있다. IaaS는 직접 VM 위에 데이터베이스를 설치하는 방식이므로 소프트웨어를 설치하거나 운영하는 기술이 있어야 운영이 가능하다. 가격 면에서는 데이터베이스를 직접 설치 및 운영하므로 PaaS에 비해 저렴(PaaS 데이터베이스 서비스 중에서는 사용하는 용량에 맞춰 종량제로 과금하는 방식도 있으므로 반드시 PaaS가 비싸진 않지만 대체로 그렇다는 의미이다. 데이터베이스 운영 노력까지 비용으로 산정하면 오히려 PaaS가 저렴하다)하다. 아마존 AWS에서는 RDS(Relational Database Service)라는 이름으로 구글 GCP에서는 Cloud SQL이라는 이름으로 데이터베이스 PaaS 서비스를 제공한다. IaaS 방식에서는 직접 컴퓨팅 VM에 설치를 하는 것이므로 AWS EC2나 GCP Compute 서비스로 VM을 만들어 설치하면 된다.

### 〈그림 11-36〉 구글 GCP에서 PaaS 방식 데이터베이스인 Cloud SQL을 생성하는 화면



출처: 구글 클라우드 플랫폼 홈페이지

데이터베이스 서비스는 모든 종류의 데이터베이스를 포괄하는 이름이고 실제로는 Oracle, MSSQL, MySQL(MariaDB), PostgreSQL 등 상세 제품을 선택해야 한다. IT 전문가가 아니더라도 일부는 한 번쯤 들어봤을 것이다. 각각에 대해 대략 설명하면 다음과 같다.

#### 〈표 11-7〉 데이터베이스 제품 비교

데이터베이스 제품	가격	설명
Oracle	높음	상용 데이터베이스 중에서 가장 유명한 제품이다. 고성능과 고급 기능이 장점이지만 높은 가격이 단점이다. 오픈소스가 없던 시절에는 절대적으로 많이 사용했지만, 높은 가격으로 인해 신규 시스템을 구축하는 경우에는 점점 사용 빈도가 줄고 있다.
MSSQL	중간	상용 데이터베이스 중의 하나로 Window 운영체제에서만 동작한다. 가격이나 기능 면에서 Oracle에 필적하지만 Window에서만 사용 가능하다는 단점이 있다.
MySQL(MariaDB)	오픈소스	신규로 구축되는 시스템 중에서 Oracle, MSSQL이 익숙해서 특별히 필요한 경우가 아니라면 가장 많이 선택되는 오픈소스 데이터베이스이다. 기본적으로 무료이고 사용자도 많아 최근에는 성능과 기능도 점점 좋아져 현재는 상용 데이터베이스에 뒤지지 않는다. 신규 서비스를 구축하는 경우라면 강력히 추천한다.
PostgreSQL	오픈소스	오픈소스 데이터베이스로 기본적으로 무료이다. 특히 Oracle 데이터베이스와 호환성이 좋아 Oracle 데이터베이스의 대안으로 많이 사용한다.

데이터베이스는 보통 SQL(Structured Query Language)이라는 언어를 통해 데이터를 처리할 수 있다. SQL은 Java나 C언어와 같은 프로그래밍 언어의 일종이지만 그것들보다는 훨씬 쉽고 직관적인 문법을 가지고 있는 언어이므로 높은 IT 지식이 없어도 사용이 가능하다. 과거에는 SQL이 IT 전문가들의 언어라고 생각되었지만 요즘처럼 IT 기술과 현업전문가의 경계가 무너지고 있는 시대에는 SQL은 공통의 언어로 생각하고 익히는 것이 좋다. 마치 영어 문법은 방대하지만 식당에서 필요한 영어가 많지 않은 것처럼 필요한 문법 몇 개만 알아둔다면 매우 유용하다. 아래는 간단한 SQL 문법을 설명한 나무위키의 내용이다.

<http://namu.wiki/w/SQL/문법>

데이터베이스에서 데이터를 저장하는 저장소 단위로 테이블이라는 개념을 사용한다. 예를 들어 회원 정보는 Users에 저장하고 구매 이력은 Transaction 테이블에 저장하면 두 테이블은 서로 다른 저장 구조를 가지고 분리되어 있다. 엑셀이나 구글시트의 시트(또는 탭)에 해당된다고 보면 쉬울 듯 하다. 아래 그림과 같이 Users, Transaction 시트가 각각 다른 컬럼을 가진 데이터를 가지고 있다. 각각이 데이터베이스의 테이블에 해당한다고 보면 된다.

〈그림 II-37〉 Users 테이블 (엑셀의 시트 또는 탭과 유사하다)

	A	B	C	D
1	Name	Gender	Address	Telephone
2	홍길동	M	서울시 종로구	111-2222-3333
3	이순신	M	전라남도 여수시	444-5555-6666
4	신사임당	F	강원도 강릉시	777-8888-9999
5				
6				
7				
8				
9				
10				
11				

만약 SQL을 사용하여 Users 테이블에서 데이터를 조회한다면 다음과 같이 된다. FROM 뒤에 있는 Users 테이블에서 Name이 '홍길동'인 레코드를 조회해서 Name, Gender, Address, Telephone 필드를 가져오는 것이다.

```
SELECT Name, Gener, Address, Telephone FROM Users WHERE Name='홍길동'
```

만약 '홍길동' 회원의 전화 번호 정보를 수정하려면 다음과 같이 한다.

```
UPDATE Telephone='111-3333-9999' WHERE Name='홍길동'
```

또 다른 테이블인 Transaction에서 가격이 500,000원 이상인 거래 레코드들만 조회한다면 어떻게 할까?

〈그림 11-38〉 Transaction 테이블

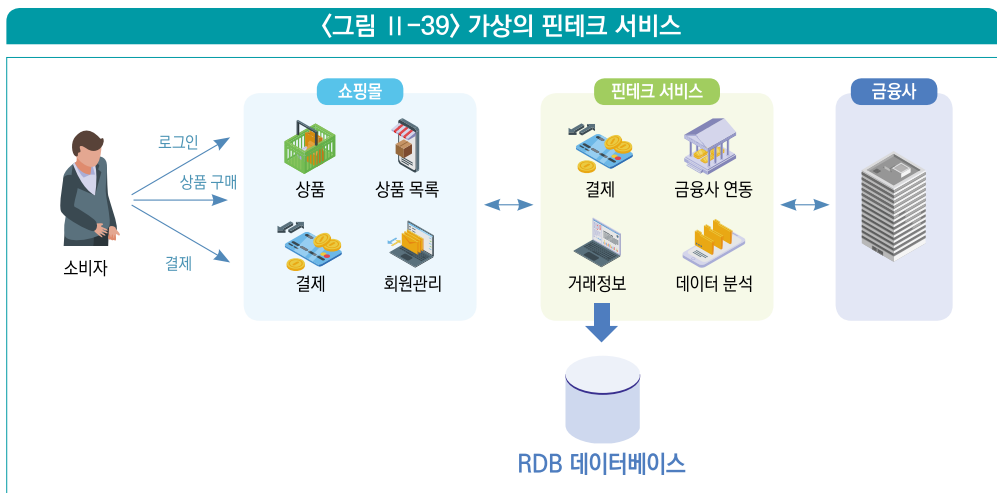
	A	B	C	D	E	F
1	Date	Item	Count	UnitPrice	TotalCost	User
2	2021-01-01	컴퓨터	10	1,000,000	10,000,000	홍길동
3	2021-01-02	컴퓨터	1	1,000,000	1,000,000	이순신
4	2021-01-03	모니터	5	200,000	1,000,000	홍길동
5	2021-01-04	모니터	2	200,000	400,000	신사임당
6	2021-01-05	키보드	20	5,000	100,000	홍길동
7	2021-01-06	키보드	2	5,000	10,000	신사임당
8	2021-01-07	책상	1	500,000	500,000	이순신
9	2021-01-08	컴퓨터	3	1,000,000	3,000,000	홍길동
10						
11						
12						

다음과 같이 하면 된다. SELECT 다음에 오는 별표(\*)는 모든 필드를 의미한다. WHERE 다음에 오는 조건이 바로 조회하고자 하는 레코드들의 규칙이다.

```
SELECT * FROM Transaction WHERE TotalCost >= 500000
```

간략하게 데이터베이스의 개념과 SQL 언어 그리고 데이터를 저장하는 테이블에 대해 알아보았다. 데이터베이스는 일반적으로 정형 데이터 저장에 유리하지만 일부 데이터베이스 중에서는 비정형 데이터 효율적인 데이터베이스도 있다. NoSQL이라고 불리는 데이터베이스들이다. 참고로 앞에서 설명한 정형 데이터 저장에 유리한 일반적인 데이터베이스는 RDB(Relational Database)이라고 한다. NoSQL 데이터베이스는 RDB에 비해 비정형 데이터를 쉽게 저장할 수 있고 RDB보다 훨씬 많은 양의 데이터를 저장할 수 있다는 장점이 있다. 클라우드 스토리지의 특징과 유사하지만 RDB처럼 조회가 빠른 게 특징이다. 반면, RDB처럼 쉽게 사용하기에는 어려움이 있어 IT 전문가와 상의하여 결정하는 좋다.

가상의 핀테크 서비스에서는 어느 업무에 어떤 데이터베이스를 사용하면 좋을까? 가상의 핀테크 서비스를 통해 일어나는 거래 정보를 저장할 때 데이터베이스 서비스를 사용하는 것이 좋고 정형화된 데이터이므로 RDB를 사용하면 된다. 아래 <그림 11-39>에는 빠져 있지만 회원 정보, 고객 상담 이력 등을 관리한다면 역시 데이터베이스에 저장해야 할 것이다. 고객 상담 이력 같은 경우에는 상담 내역이 길고 유형이 다양해서 NoSQL로 볼 수도 있지만, 사실 웬만한 데이터는 RDB에서 처리가 가능하고 상대적으로 적은 IT 지식으로도 사용이 가능하므로 RDB를 추천한다. RDB 제품 중에서는 MySQL을 권고하고 PaaS 형태로 사용하기를 권장한다. 이유는 가격이 IaaS 방식에 비해 다소 비싸더라도 훨씬 사용이 쉬워 관리에 들어가는 인력을 줄일 수 있어 전체적인 투자 비용은 오히려 저렴하다.

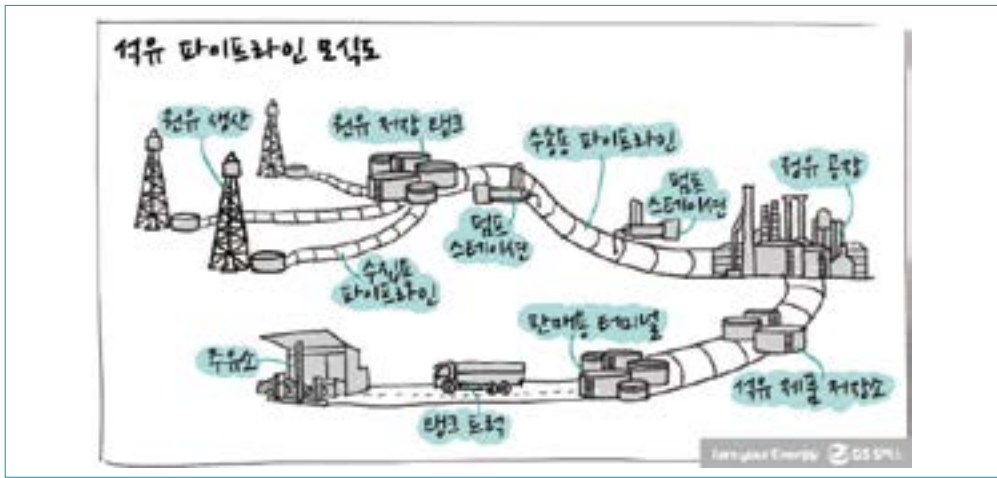


## 5 빅데이터 분석

빅데이터 또는 데이터 분석은 핀테크에서 중요한 부분이자 클라우드의 장점을 극대화할 수 있는 영역이다. 빅데이터 중요성은 빅데이터와 인공지능 챗터에서 자세히 확인할 수 있다. 빅데이터는 대량의 데이터를 수집 및 분석하여 비즈니스 업무나 서비스에 필요한 인사이트를 도출하는 과정이다. 빅데이터는 이름에서도 알 수 있듯이 정말 많은 대용량의 데이터를 다루어야 하는 부분이다. 많은 양의 데이터를 다루기 위해서는 우선 시스템의 규모가 엄청나게 커야 하며 클라우드는 대규모 인프라를 효율적으로 제공할 수 있는 서비스이므로 빅데이터와 클라우드는 찰떡 궁합이라고 할 수 있다. 또한 빅데이터는 대량의 데이터가 잘 정제된 형태가 아닌 경우가 많아 가공 과정이 필요하고 이때 여러 가지 소프트웨어들과 시스템이 부가적으로 필요하다. 그것은 곧 해당 시스템과 소프트웨어를 설치 운영하기 위한 기술과 인력 그리고 비용이 필요하다는 의미이다. 클라우드에서는 이러한 시스템과 소프트웨어들을 PaaS 형태로 제공하기 때문에 직접 설치 및 운영할 필요가 없다. 비용면에서도 필요한 부분만 가져다 사용할 수 있어 직접 하드웨어를 구매해서 구축하는 것에 비해 효율적이다.

데이터는 21세기의 원유라고 한다. 원유만큼이나 중요한 자원이라는 의미이다. 그런데 재미 있는 점은 원유를 채굴해서 실제 자동차의 휘발유로 넣는 과정과 데이터를 수집에서 분석하는 과정은 상당히 유사하다. 다음 <그림 II-40>은 원유를 생산해서 정제 후 휘발유로 바뀌어 각 주유소로 가는 과정을 나타낸 그림이다. 알다시피 유정에서 바로 뽑아올린 원유는 그대로 사용할 수 없다. 정유공장에서 정제 과정을 거치며 불순물을 제거하고 원하는 유종(휘발유, 등유 등)을 추출해야 사용할 수 있다. 일반적으로 정유공장은 원유 산지에 바로 붙어 있지 않기 때문에 원유를 수송하는 과정이 필요하다. 추출된 휘발유나 등유 등은 저장소에 보관되었다가 실제 그것을 필요로 하는 가정이나 주유소로 배송된다.

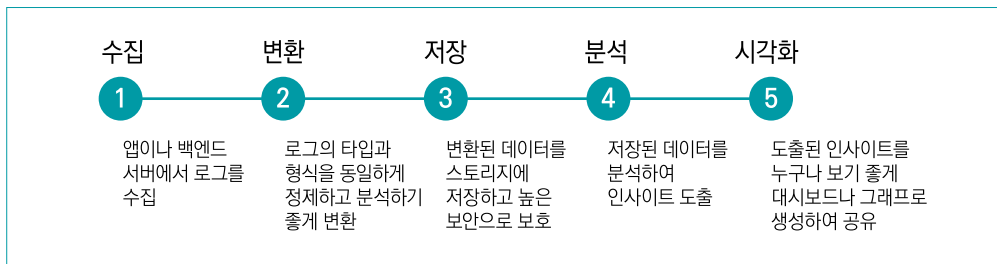
〈그림 11-40〉 원유 정제 과정



출처: GS 칼텍스 미디어허브

아래 〈그림 11-41〉은 일반적인 데이터 분석 작업을 나타낸 것이다.

〈그림 11-41〉 데이터 분석 과정

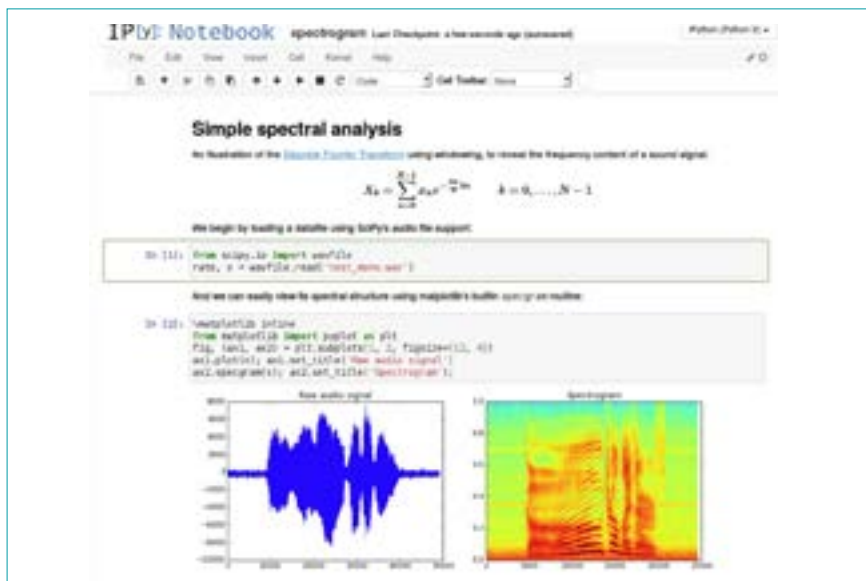


빅데이터 분석 과정은 이러한 원유 정제 과정과 상당히 유사하다. 데이터도 원유의 유정과 마찬가지로 여러 가지 데이터를 만들어지는 시스템에서 생산된다. 예를 들어 스마트폰의 데이터는 스마트폰에서 온라인 쇼핑몰의 고객 행동 데이터는 쇼핑몰 웹서버에서 생성된다. 생산된 데이터는 그곳에서 바로 분석이 이루어지기보다는 분석을 위한 소위 빅데이터 시스템으로 전송이 되어야 한다. 이렇게 수집된 데이터는 1차적으로 불필요하거나 오류가 있는 데이터를 제거하고 분석이 편리한 형태로 변환어야 한다. 정유공장에 비유될 수 있다. 실제로 데이터 분석을 해 보면 수집된 데이터에는 오류도 많고 분석에 최적화된



원본(raw) 데이터가 아니라 분석이 용이하지 않다. 많은 경우 원본 데이터는 비정형인 경우가 많은데 그것을 직접 분석하기는 쉽지 않아 정형 데이터로 변환하는 경우가 많다. 수집된 데이터는 대규모 저장소에 저장되는데 이때 사용되는 저장소는 클라우드 스토리지나 데이터베이스이다. 앞에서 설명한 것처럼 대규모 비정형 데이터는 클라우드 스토리지에 저장하는 것이 유리한 반면 정형 데이터로 변환된 데이터는 데이터베이스에 저장하는 것이 유리하다. 원유 저장소에 비유될 수 있다. 원유의 저장소에는 1차로 생산된 원유 저장소가 있고 정유가 끝난 제품의 저장소가 있듯이, 데이터의 저장소도 반드시 변환 후나 분석 전에 보관해야 하는 것은 정해진 규칙은 아니다. 상황에 맞게 어떤 경우는 변환 전에 원본 데이터를 대규모로 저장하기도 하고 분석이 끝난 데이터를 다시 저장소에 보관하기도 한다. 데이터 분석가들은 저장된 데이터를 이용하여 분석 작업을 수행한다. 분석을 거쳐 수백 테라바이트 데이터에서 필요한 인사이트를 가진 수십 메가바이트 데이터가 나오기도 한다. 이것은 마치 정유공장을 거쳐 휘발유나 등유처럼 각각의 유종이 만들어지는 것과 유사하다. 추출된 데이터는 일반적으로 시각화(차트나 표 등으로 알아보기 쉬운 리포트로 변환) 과정을 통해 현업 전문가나 의사결정자에게 전달이 된다. 이 부분은 원유에서 추출된 휘발유가 주유소로 전달되어 각 차량에 들어가는 것과 유사하다고 볼 수 있다.

〈그림 11-42〉 Jupyter Notebook 분석 도구

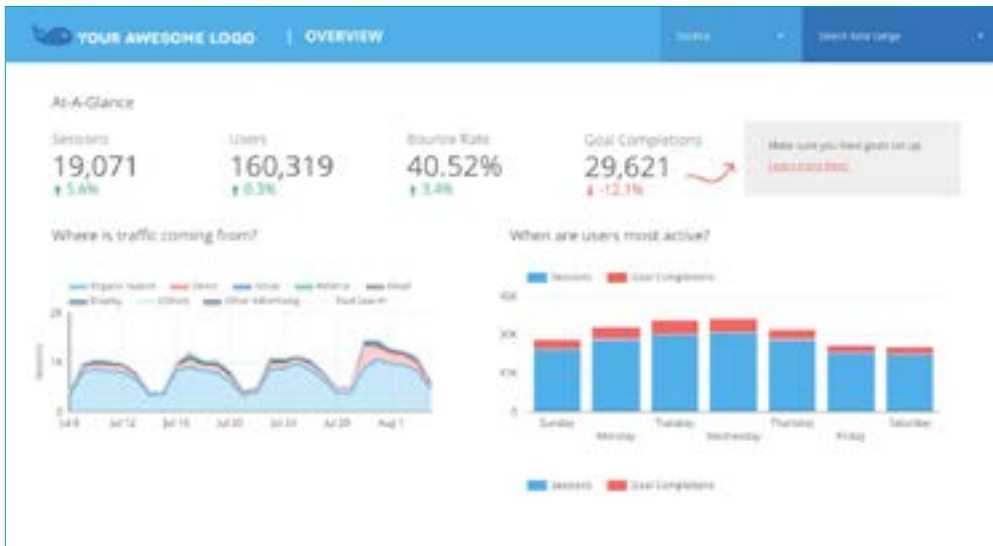


출처: <https://arogozhnikov.github.io/images/jupyter/example-notebook.png>

당장 보아도 빅데이터 분석이라는 것이 단번에 이뤄지는 과정이 아니란 것을 알 수 있다. 각 과정에 여러 가지 소프트웨어와 시스템이 필요한데 클라우드에는 여기에 맞는 서비스들을 제공한다. 데이터 수집 같은 경우에는 구글 애널리틱스 같은 서비스를 통해 웹사이트 고객의 행동 데이터를 추출하고, CDC/ETL 등의 서비스를 통해 다른 업무시스템의 데이터를 추출할 수 있다. 큐(queue)를 통해서 추출된 데이터를 중앙 저장소로 마치 원유 파이프라인처럼 보낼 수 있다. 변환 과정에는 역시 CDC/ETL 서비스(원본 데이터 소스에서 원하는 대상으로 데이터를 옮겨 주는 기능)를 사용하기도 하고, Spark(Apache 재단에서 관리하는 오픈 소스)나 Dataflow(구글 GCP가 제공하는 Spark와 유사한 서비스. 타 클라우드에서는 Spark 기반 서비스를 주로 활용) 같은 서비스를 통해 데이터를 가공하고 비정형 데이터를 정형 데이터로 변환한다.

저장소는 위에서 말한 것처럼 클라우드의 클라우드 저장소나 데이터베이스 서비스를 이용한다. 아마존 AWS의 경우에는 Redshift나 구글 GCP의 경우에는 BigQuery와 같은 빅데이터 전용 저장소 서비스를 제공하기도 한다. 분석을 위해서는 데이터 분석가들을 위한 노트북(Jupyter Notebook) 서비스나 Spark, Dataflow 서비스들을 활용한다. 특히 구글에서 무료로 제공하는 colab(<http://colab.research.com>) 같은 서비스는 비용이 들지 않고 여러 사람들이 협업하기에 편리하므로 많이 사용된다. 시각화는 주로 BI(Business Intelligence)를 많이 사용한다. 시각화를 위한 BI 도구는 사용자의 수준이나 용도에 맞게 다양한 제품들이 사용되는데 클라우드에서는 무료부터 상용까지 다양하게 제공하고 있다.

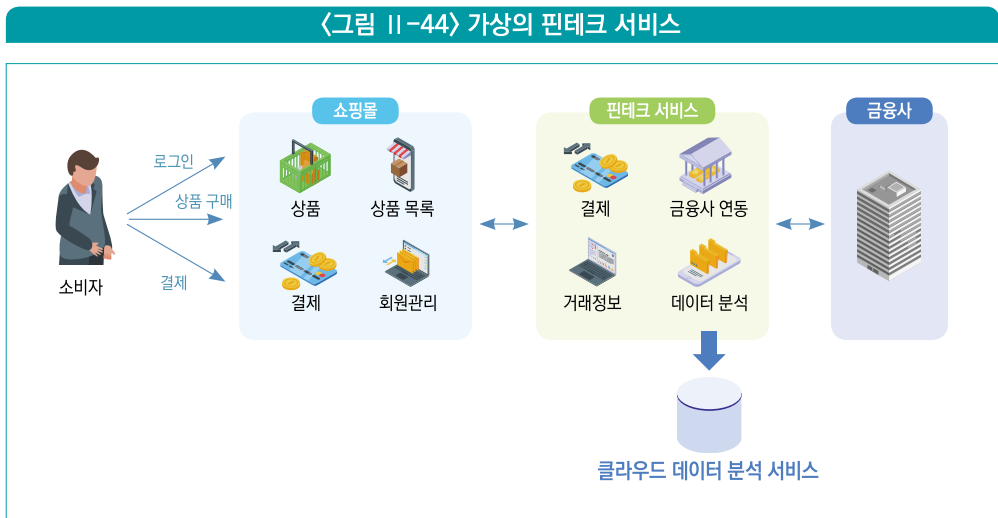
〈그림 11-43〉 구글이 제공하는 무료 데이터 시각화 BI 도구인 Data Studio



출처: <https://kasatria.vn/wp-content/uploads/2019/04/google-data-studio-img.png>

지금까지 클라우드를 이용한 빅데이터 분석 방법에 대해 간략하게 설명하였다. 클라우드가 데이터 분석을 위한 많은 서비스들을 제공하고 있어 쉽게 분석 작업을 완료할 수 있다는 것도 알았다. 다만, 데이터 분석은 도구도 중요하지만 비즈니스 요구사항에 대한 이해와 그에 맞는 분석 경험이 매우 중요하다. 실제 빅데이터 분석을 통해 어떤 비즈니스 문제를 풀고 싶은지 정의하는 것이 우선이다. 데이터 분석은 목표가 분명해야 성공할 수 있다. 그리고 나서 그것을 어떻게 풀지 고민을 해야 하는데 이 과정은 데이터 전문가와 상의가 필요한 영역이다. 물론 전문가에게 맡긴다고 100% 되는 것은 아니고 비즈니스 오퍼로서 업무 경험과 인사이트를 제공하는 것은 너무나 중요하다. 같은 분석이라 하더라도 어떤 목표와 어떤 철학을 가지고 하는지에 따라 결과는 천차만별이다. 실제 분석 과정은 시간이 많이 드는 과정이다. 목표와 방향이 분명하지 않다면 시간을 많이 쓰기도 비즈니스에 도움이 안되는 결과를 얻을 수 있다. 클라우드의 빅데이터 서비스는 사이클 선수에게 최고급형(그런데 가격도 저렴한) 사이클과 같다. 선수의 기량이 좋다면 고급 사이클을 통해 좋은 기록을 얻을 수 있을 것이고, 기량이 좋지 않다면 좋은 기록은 기대하기 어려울 것이다. 반면 기량이 좋은 선수도 기능과 상태 면에서 너무 안좋은 사이클로는 좋은 기록을 내기 어렵다. 클라우드 기반의 빅데이터 서비스는 데이터 분석가가 가진 역량을 극대화할 수 있는 도구임을 명심하자.

가상 핀테크 서비스에서는 어느 부분에 클라우드 데이터 분석 서비스를 활용할 수 있을까? 가상 핀테크 서비스에서는 고객의 거래 내역을 보고 우수한 고객에게 유리한 금리로 금융서비스를 추천한다고 비즈니스 목표를 가정해 보겠다. 이 부분에 클라우드 데이터 분석 서비스를 활용할 수 있다.



우선 데이터 분석 서비스에서는 거래 정보가 저장된 데이터베이스로부터 고객(User), 거래(Transaction) 데이터를 가져와야 할 것이다. 이 경우에는 CDC/ETL 서비스를 사용하면 된다. 고객의 소셜 데이터를 분석한다면 대상이 되는 소셜 서비스로부터 큐를 통해 데이터를 가져와 Spark나 Dataflow 등을 이용해 가공한 후 저장소에 저장해야 할 것이다. 만약 금융서비스 페이지에서 고객이 어떤 서비스에 관심을 보이는지 알고 싶다면 구글 애널리틱스 같은 서비스를 이용해서 데이터를 수집할 것이다. 이 부분은 원하는 비즈니스 목표에 따라 모습이 많이 달라지므로 전문가와 상의하기를 바란다.

## 6 AI/ML

인공지능(AI, Artificial Intelligence)와 기계학습(ML, Machine Learning)은 사실 데이터 분석의 한 분야였다. 그러나 이 분야를 통해 얻는 성과들이 많아지면서 현재는 별도의 분야로 간주되고 있다. 콜센터에서 인공지능 상담사가 인간 대신 고객들의 불만을 처리하기도 한다. 쇼핑몰에서 상품이미지가 상품과 맞는지 비전 AI를 통해서 감별하기도 한다. 소셜에 올라온 포스팅을 분석해 상품에 대한 반응을 분석하기도 한다. 제조업에서는 인공지능을 통해 수요를 예측해 필요한 만큼의 상품만 생산하여 비용을 절감한다. 데이터 분석이 비즈니스 인사이트를 얻기 위해 필요한 데이터를 수집하고 분석을 위한 환경을 제공한다면, AI/ML은 이러한 데이터를 활용하여 사람 대신 인사이트를 제공하는 역할을 한다. 사람이 하던 분석 작업이 시간이 오래 걸리고 상당히 높은 숙련도와 경험을 필요로 하는 일이다 보니 분석가의 역량에 따라 결과가 너무 들쭉날쭉했다. AI/ML은 이러한 부분을 높은 컴퓨팅 성능을 통해 해결하는 과정으로 실제 비즈니스 요구사항에 바로 적용할 수 있는 서비스를 제공할 수 있는 장점을 가진다. 클라우드는 AI/ML에 필요한 대규모 IT 시스템을 제공하거나, 더 나아가 비즈니스 요구사항에 딱 맞는 맞춤형 AI/ML 서비스를 직접 제공한다.

AI/ML은 몇 가지 영역으로 나뉜다. 각 영역별로 클라우드가 어떤 서비스를 제공하고 있는지 살펴본다.

### 6-1 비전 AI

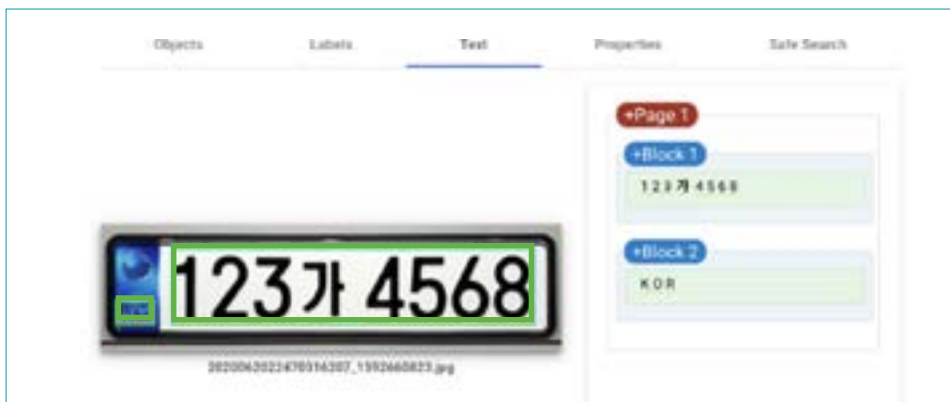
인공지능이 빠르게 인간의 능력을 따라잡아 효과를 보고 있는 분야이다. 이미지나 영상을 보고 어떤 것인지 분류하는 것이 일반적인 사용 예이다. 다음은 비전 AI를 활용한 예이다.

- 얼굴 이미지를 보고 누구인지 식별한다 (face recognition)
- 자동차 번호판을 인식한다 (OCR)
- 종이 문서를 보고 쓰여진 글자를 입력한다 (OCR)
- 자동차 부품을 찍어 해당하는 매뉴얼을 제공한다 (image recognition)
- CCTV를 보고 1시간 동안 지나간 차량의 대수와 차종을 알아낸다 (video analysis)
- 축구 중계에서 공의 위치를 식별하여 카메라를 자동으로 움직인다 (object detection)

사람은 많은 정보를 시각을 이용해 획득한다. 그것을 인공지능으로 대체하는 것이다. 얼굴 이미지를 보고 사람을 식별하는 기능의 경우에는 소셜 서비스에서 사진의 얼굴 태깅에 활용하기도 하고, 조금더 정교한 정확도를 확보하여 스마트폰의 잠금해제에 사용하기도 한다. 비전 AI는 태깅(얼굴 사진이 누구의 것인지, 영상의 몇 초부터 사람이 등장하는지, 축구 이미지에서 공의 위치 등)된 사진이나 영상과 같은 시각자료가 대량으로 있으면 비교적 쉽게 구현할 수 있다. 오히려 태깅 작업이 어려워 해당 부분을 서비스로 제공하기도 한다. 태깅은 사람이 직접해야 하는 작업이므로 클라우드 업체들은 보통 사람들을 외주로 고용하여 서비스를 만들었다. 태깅된 이미지들을 저장하고 관리하는 작업도 손이 많이 가는 작업인데 이런 부분들을 서비스로 제공하기도 한다.

클라우드가 제공하는 주요 비전 AI 서비스로는 Vision API와 AutoML이 있다. Vision API는 미리 만들어진 인공지능 모델로서 특별한 학습 과정이 없이 바로 적용이 가능하다. 자동차 번호판 같은 경우는 대부분의 이미지가 유사하기 때문에 별도로 학습할 필요가 없다. 이런 경우는 단지 해당 서비스를 API로 호출만 하면 된다. 아래 <그림 11-45>는 실제 번호판을 구글 GCP의 Vision API 데모 페이지에서 인식해 본 것이다. 번호판의 글자는 비교적 단순하기 때문에 어렵지 않게 인식이 된다.

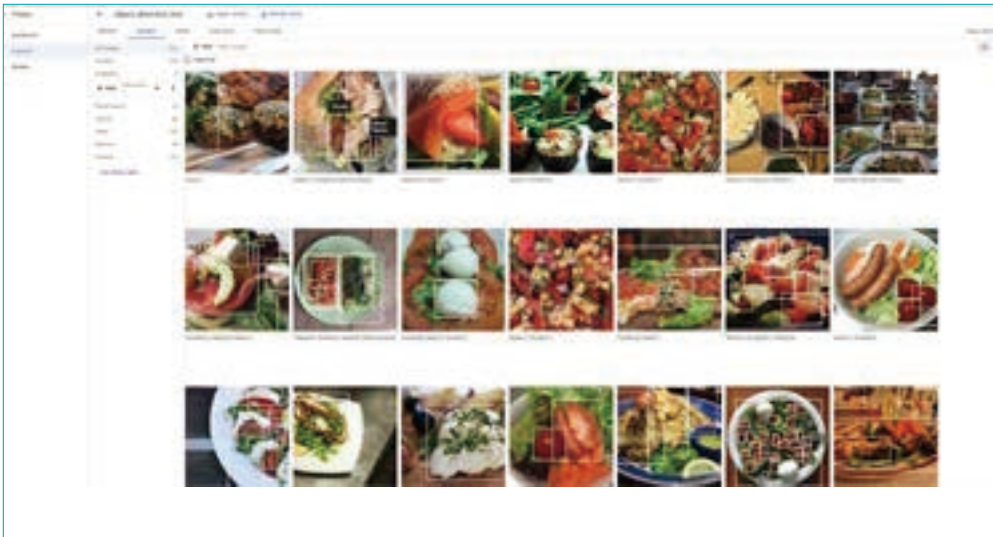
<그림 11-45> Vision API를 통한 번호판 인식(예시)



출처: Google vision API

만약 누군가 손으로 쓴 글씨를 인식해야 한다면 Vision API는 어떨까? 워낙 사람마다 글씨체가 다르고 흘려 쓰는 경우도 많아 인식이 쉽지는 않다. 이런 경우는 AutoML을 활용할 수 있다. AutoML은 태깅된 이미지가 있다면 인공지능 모델을 자동으로 만들어 주는 기능이다. 그러면 Vision API처럼 매우 정형화된 공통의 이미지뿐만 아니라 특수한 이미지도 식별할 수 있다. 대표적으로 사람의 얼굴을 대상으로 Vision API는 단지 사람이라고만 식별하는데, AutoML을 이용하면 특정 인물의 사진인지 구체적으로 식별할 수 있다.

〈그림 11-46〉 AutoML을 이용한 비전 AI 모델 생성



출처: Google Cloud AutoML Vision

## 6-2 음성인식(Speech to Text) 및 음성발화(Text to Speech)

비전 다음으로 많이 인공지능이 활용되는 분야는 음성이다. 유튜브 영상을 보면 자동으로 자막이 생성되는 기능을 유용하게 사용하고 있을 것이다. 이것은 음성을 인식하여 텍스트로 변환해 주는 STT(Speech to Text) 기능을 활용한 것이다. 반대로 텍스트를 입력하면 음성으로 만들어주는 기능도 있는데 TTS(Text to Speech)라고 한다.

<그림 11-47> 유튜브 자막 서비스



출처: 구글 유튜브 비디오

STT나 TTS의 경우에도 클라우드에서는 비전 AI와 마찬가지로 직접 모델을 만들 필요 없이 API 형태로 바로 사용할 수 있는 기능을 제공한다. 아래 웹사이트에서 STT, TTS 기능을 직접 테스트해 볼 수 있다. 인공지능 기반 콜센터의 고객 응대는 STT를 이용해 고객의 목소리를 인식하고 TTS를 통해 응답을 내보는 것이다.

- STT 데모: <https://cloud.google.com/speech-to-text>
- TTS 데모: <https://cloud.google.com/text-to-speech>

### 6-3 자연어 처리(NLP, Natural Language Processing)

가장 활용도가 높은 분야이고 최근에 많은 발전을 이루어진 분야이다. 소셜 네트워크에 올라온 고객의 피드백을 분석해 제품에 대한 반응이 긍정적인지 부정적인지 알려진 자연어 처리가 반드시 필요하다. 챗봇을 만드는 경우에도 자연어 처리는 필수이다. 자연어의 활용 사례는 다음과 같다.



- 텍스트의 내용이 긍정적인지 부정적인지 판단 (sentiment analysis)
- 텍스트에서 의미가 있는 항목 추출 (entity extraction)
- 뉴스 기사를 몇 줄의 문장으로 요약 (summarization)
- 뉴스 기사를 분야 별로 자동 분류 (text classification)
- 콜센터나 챗봇에서 고객의 질문에 대해 답변 (text classification, entity extraction 등)

자연어 처리 인공지능은 클라우드에서 여러 가지 형태로 지원 받을 수 있다. 첫 번째는 NLP API를 활용하여 별도의 모델 없이 사용하는 방법이다. 비전, 음성과 마찬가지로 API 호출을 통해 복잡한 과정 없이 바로 원하는 기능을 얻을 수 있다. 다른 API 기반 인공지능 서비스와 마찬가지로 공통적인 데이터를 기반으로 학습되어 있기 때문에 특별한 목적보다는 범용 목적의 텍스트 분석에 적합하다.

- NLP 데모: <https://cloud.google.com/natural-language>

두 번째는 AutoML을 이용하는 방법이다. 비전 AutoML과 마찬가지로 태깅된 텍스트 데이터가 있다면 모델을 자동으로 만들어 준다. 자연어 처리는 세부 분야가 매우 다양한데 AutoML을 통해서도 그중에서 분류(text classification), 항목 추출(entity extraction), 감정 분석(sentiment analysis) 등의 모델을 만들 수 있다.

또 한 가지 클라우드의 자연어 처리 인공지능 서비스는 챗봇이다. 챗봇은 여러 가지 인공지능이 복합적으로 구현된 것으로 인간과 인공지능이 대화를 하는 데 초점을 맞춘 서비스이다. 사람의 대화에서 의도(intent)를 파악하고 그에 맞는 답변을 제공할 수 있도록 설계되어 있다. 어떤 대화를 진행할 것인지는 챗봇에 알려주어 학습이 되도록 해야 한다.

자연어 처리 인공지능은 분야가 다양하여 클라우드에서 기본적으로 제공되는 기능만으로 모두 커버할 수는 없다. 예를 들어 텍스트를 요약하는 기능 같은 경우에는 분야별(의료 분야 vs IT 분야)로 워낙 데이터가 달라 직접 해당 분야의 데이터를 가지고 학습을 하지 않으면 안 된다. 이런 경우에는 알려진 유명한 인공지능 모델을 가져와 추가 학습을 진행해야 한다. 클라우드에서는 모델 카탈로그로 원하는 모델을 조회하고 다운받을 수 있는 서비스를 제공하고, 다운받은 모델을 학습할 수 있는 인공지능 학습용 서비스(주로 GPU 기반)를 제공한다.

## 6-4 정형 데이터 분석

핀테크 서비스를 만들 때 아마도 가장 유용한 분야일 것이다. 앞에서 설명한 비전, 음성, 자연어처리는 회원정보나 거래정보 등과 달리 엑셀이나 데이터베이스 테이블의 형태로 저장하기 어려운 비정형 데이터에 해당한다. 하지만 예를 들어 구매 이력을 바탕으로 새로운 상품을 추천하는 경우에는 대부분 정형 데이터를 가지고 인공지능 모델을 만들게 된다. 정형 데이터 분석의 예와 필요한 데이터는 다음과 같다. 여러분들이 하고자하는 비즈니스 모델과 요구사항을 바탕으로 머리속에 그려보기 바란다.

- 상품 추천: 상품 조회 또는 구매 이력, 회원 정보 데이터
- 사기 탐지(fraud detection): 거래정보, 상품정보 데이터
- 수요 예측: 거래정보, 상품정보, 매장정보 데이터
- 고객 생애 가치(lifetime value) 예측: 구매 이력, 회원 정보 데이터

정형 데이터 분석에서 만들어지는 인공지능 모델은 대략 다음과 같다. 데이터 분석이나 AI/ML에 익숙하지 않은 분들이라면 아래 인공지능 모델에 다소 어려운 개념일 것이다. 위에서 말한 적용 사례는 비즈니스에 가까운 개념이지만, 아래는 수학 모델에 가까운 기술적 개념이기 때문이다. 따라서 이 부분에 대한 자세한 내용은 빅데이터와 인공지능 챕터를 참고하기 바라고 정형 데이터 분석을 이해하기 위한 용어라고 알아두면 좋겠다.

- 회귀(regression): 유사한 성향을 가지는 고객이나 상품의 데이터를 바탕으로 구매 액수, 상품 수요 등을 예측. 주로 수요 예측, 고객 생애 가치 예측 등에 사용
- 분류(classification): 유사한 성향을 가지는 데이터를 바탕으로 고객이 회원 가입할지, 구매 여부 등을 예측. 주로 타겟 마케팅 등에 사용
- 오토인코딩(auto encoding): 복잡한 데이터의 기본적인 성질만 추출하거나 또는 그것을 바탕으로 다시 복원하여 이상(anomaly) 여부를 예측. 주로 사기 탐지, 이상 감지 등에 사용
- 클러스터링: 태깅된 데이터 없이(비지도 학습) 단지 데이터의 성질만 가지고 카테고리(클러스터)를 나눠 분류. 주로 타겟 마케팅, 이상(anomaly) 감지 등에 사용

클라우드에서는 위에서 말한 비즈니스 사례에 대해서 인공지능 모델을 만들기 위한 서비스들을 제공한다. 대표적으로는 AutoML 기술을 활용하여 상품 추천이나 수요 예측과 같은 모델을 미리 만들어 놓고 데이터만 입력하면 되도록 서비스를 제공하고 있다. 직접 모델을 만드는 것에 비해 상당히 쉬우면서도 최신 모델을 활용할 수 있는 장점이 있다. 따라서 원하는 비즈니스 사례를 바탕으로 클라우드에서 제공하는 빌트인(built-in) 서비스들을 먼저 검토해 보기 바라고 맞는 서비스가 없다면 직접 모델을 만드는 것을 고려하기를 권고한다.

〈그림 11-48〉 아마존 AWS에서 제공하는 수요 예측 모델 서비스



출처: 아마존 AWS 홈페이지

만약 원하는 서비스가 없다면 직접 모델을 만들어야 한다. 클라우드가 제공하는 인공지능 학습용 서비스를 이용하기 바란다. 클라우드 별로 대략 아래와 같은 서비스들을 제공하고 있다.

- 아마존 AWS의 SageMaker
- 구글 GCP의 Vertex AI(구글 Cloud ML 엔진)
- 마이크로소프트 Azure의 AI 플랫폼

가상 핀테크 서비스에서 활용할 클라우드의 AI/ML 서비스는 실제 구체적인 비즈니스 프로세스가 정해지지 않아 단적으로 말하기는 어렵겠지만, 대략 다음과 같은 서비스들을 고려해 볼 수 있을 듯 하다.

- 비전 AI: 고객의 신분증을 받아 글씨를 읽어 입력하는 기능
- 음성인식: 인공지능 콜센터 또는 상담센터
- 자연어 처리: 고객의 댓글에서 심각한 사안에 대한 신속 대응
- 정형 데이터 분석: 금융상품 추천, 고객 세그멘테이션을 통한 타겟 마케팅

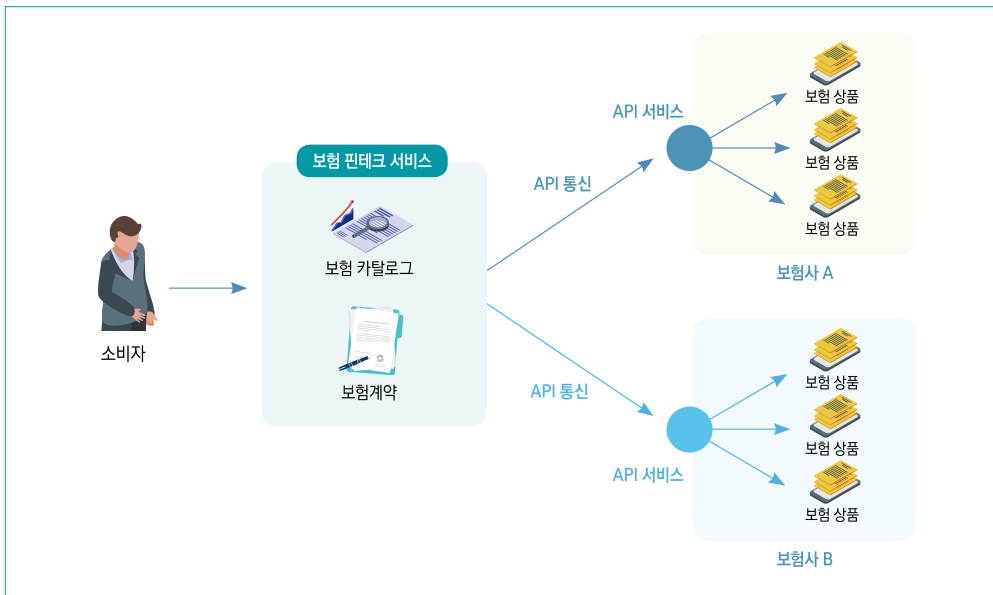
## 7 API 서비스

이 글에서 여러 번 API(Application Programming Interface)에 대해서 언급했다. 사람은 말(한국어, 영어)을 가지고 대화를 한다. 컴퓨터 시스템들은 일정한 기술적 프로토콜(기계가 인식하는 언어에 해당)을 가지고 통신한다. 따라서 API란 쉽게 말하면 시스템 간에 또는 프로그램이 의사를 서비스에 전달할 때 사용하는 의사소통 도구(interface)라고 생각하면 된다. 핀테크에서는 이러한 API가 매우 중요한 역할을 한다. 많은 핀테크 서비스가 스스로 금융상품을 제공하는 경우 보다는 이미 시장에 존재하는 금융상품을 소비자에게 편리하게 제공하고 있다. 예를 들어 보험사들의 상품을 모아서 고객에게 제공하는 경우에는 보험사들과 통신하여 상품들을 가지고 와야 한다. 결제 서비스의 경우에는 카드사나 은행과

협업이 필요하다. 과거라면 전화나 서류를 통해 이러한 과정을 진행하겠지만, 금융과 IT 시스템이 결합된 핀테크 서비스는 시스템 간 통신을 통해 온라인 방식으로 처리한다. 이때 주로 사용하는 것이 바로 API이다.

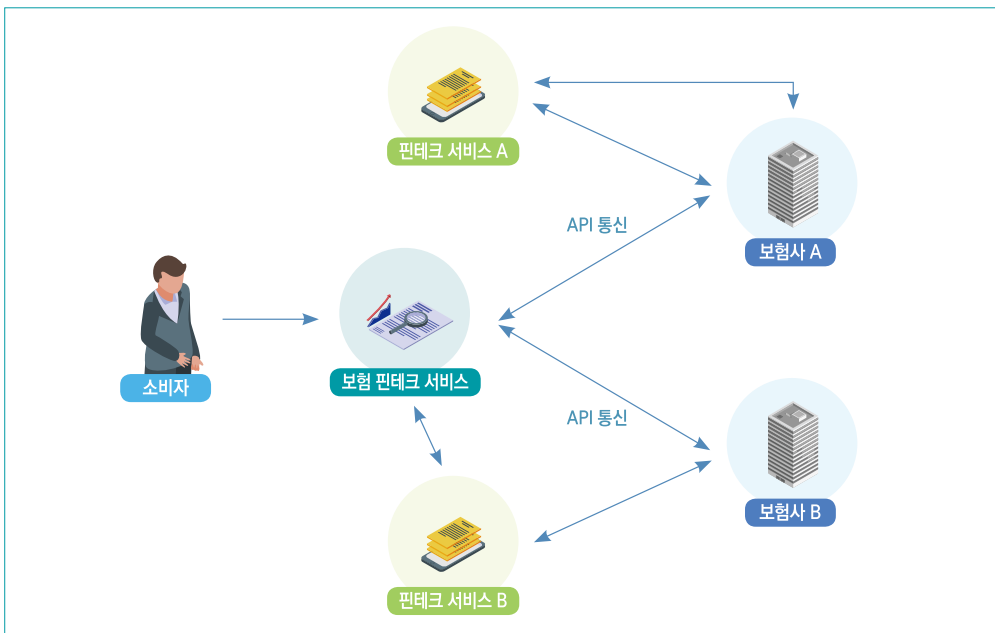
아래 <그림 11-49>는 가상의 보험 핀테크 서비스에서 API를 활용하여 실제 금융상품을 제공하는 보험사와 통신하는 모습을 나타낸 것이다. 보험사 A, B는 각각 다양한 보험 상품들을 개발할 것이다. 소비자에게 여러 보험사의 상품들을 비교, 추천하는 이 서비스에서는 두 보험사의 최신 상품 정보를 가져와야 한다. 때로는 소비자의 보험 계약을 대행해야 할 수도 있다. 이때 보험사는 자사의 보험 상품을 조회하고 계약을 진행할 수 있는 기능을 API로 만들어 핀테크에 서비스에 제공한다. 그러면 핀테크 서비스는 보험사가 제공하는 API 규약(사람들 간에 대화에도 예절이 존재 하듯이 미리 정해진 약속)을 이용하여 호출하는 프로그램을 작성한다. 소비자 입장에서는 개별 보험사 시스템에 접속하지 않아도 금융사와 API 통신을 구현한 핀테크 서비스를 통해 모든 보험 상품을 조회하거나 계약을 진행할 수 있다.

<그림 11-49> 가상의 보험 핀테크에서 API 활용 예



API는 시스템 간 혹은 회사 간에 서비스를 연동하는 효과를 낼 수 있는 기술이다. 이러한 서비스 모델은 핀테크 생태계를 만드는 역할을 한다. 보험사가 만든 API는 다른 핀테크 서비스에서도 불러 사용할 수 있다. 핀테크 업체 또한 다른 핀테크나 금융사에 제공하는 API를 만들어 자사의 서비스를 제공할 수도 있다. 마치 여러 나라가 자국의 항구를 무역항으로 열면 그것을 통해 통상이 진행되는 것과 비슷하다. API는 자사의 서비스를 무역항으로 개방하는 것과 유사한 역할을 한다.

〈그림 11-50〉 API를 통한 핀테크 생태계 확장



각 클라우드에는 이러한 API 서비스를 쉽게 구현할 수 있는 기능들을 제공하고 있다. 또는 클라우드 위에 API 기능을 제공하는 제품들을 얹어 사용할 수도 있다. API 분야에서 유명한 솔루션으로는 구글 GCP의 apigee와 브로드컴의 Layer 7 제품이 있다. apigee는 구글 GCP에서 바로 클라우드 형태로 제공하고 있는 서비스이고 Layer 7은 설치형 제품이라 필요하면 아마존 AWS나 마이크로소프트 Azure 등에 구축하여 사용할 수 있다. 이러한 제품들을 통칭해서 API Gateway라고 한다. AWS와 Azure에서는 자체 API 서비스(각각 Amazon API Gateway, Azure API)를 제공하기도 한다.

가상의 핀테크 서비스에서는 금융사와 연동하는 부분에 API Gateway 서비스를 두어야 한다. 물론 여기서는 주로 핀테크 서비스에서 금융사를 호출하는 입장(호출받는 쪽에서 API 서비스를 구현한다)이기 때문에, 사실은 금융사에서 자사의 핀테크 서비스를 위한 API를 구현하는 게 맞다. 핀테크 서비스 입장에서는 자사의 서비스를 API로 만들어 핀테크 생태계에 참여하고 싶다면 역시 API Gateway를 이용하여 서비스를 개방해야 한다.

〈그림 11-51〉 가상의 핀테크 서비스



## 8 IoT

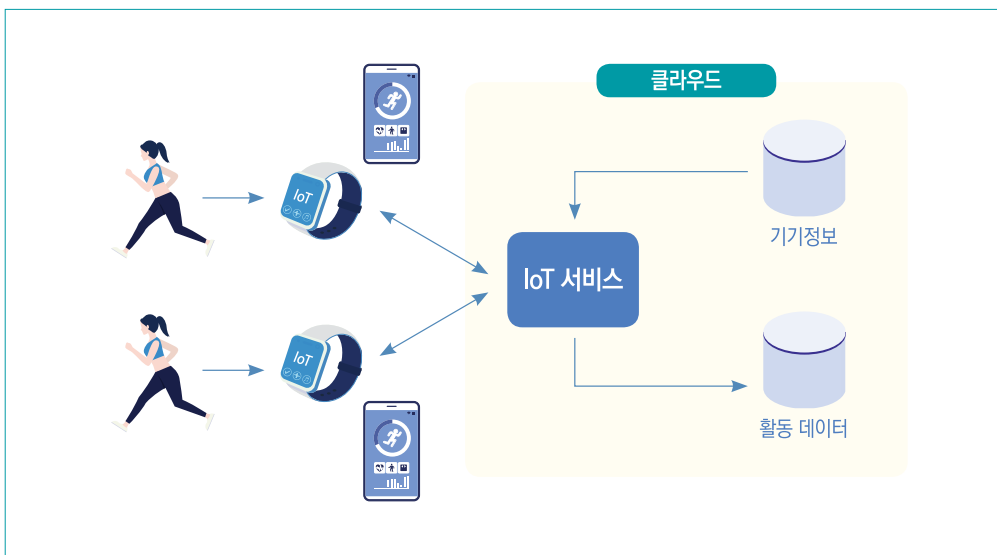
IoT는 Internet of Things의 약자이며 우리 말로는 사물인터넷이라고 한다. IoT를 쉽게 정의하면 우리가 인터넷을 통해서 전세계 각국의 개인, 기업들과 통신(소셜 메시지, 비즈니스 거래)을 하듯이 기기들, 그중에서도 주로 모바일 기기들이 서로 통신을 하는 것이다. 스포츠 시계, 전기 자동차, 스마트폰 결제 등이 IoT 통신의 대표적인 사례이다. IoT 기기들의 특징은 통신 기능이 있다는 것인데 실제로는 그 안에 일종의 컴퓨터가 들어가 연산도 가능하다는 점이다. 모든 사물에 컴퓨터와 통신 기능이 들어가는 것이 IoT라고 생각할 수 있다. 자세한 내용은 IoT 챕터를 참고하기 바란다.

핀테크 분야에서 IoT 기술은 대표적으로 모바일 결제에 사용될 수 있다. 요즘 신용카드를 가지고 다니지 않아도 스마트폰 하나로 결제가 가능한 시대이다. 이때, 신용카드 역할을 하는

스마트폰과 결제를 수행하는 POS(Point Of Sales) 단말은 서로 IoT 통신(IoT 통신이 특별히 기존 기술과 다른 것은 아니지만, 중앙의 컴퓨터를 거치지 않고 직접 기기 간에 통신한다는 의미에서)을 한다. 분야를 조금 더 확장해 보면 스포츠 시계를 착용하는 사람들을 대상으로 운동 기록을 확보해 보험 비용을 다르게 적용할 수도 있다. 자동차 보험사들은 이미 자동차 운행 기록을 확보하여 자동차 보험 갱신에 활용하고 있다.

스포츠 시계를 기반으로 한 IoT 서비스를 가정하고 클라우드의 역할에 대해서 알아보자. 스포츠 시계는 주로 착용자의 신체적인 활동 내용을 수집하는 역할을 한다. 수집된 정보는 직접(자체적으로 4g/5g 무선통신 기능이 있는 경우) 또는 스마트폰을 거쳐 클라우드에 있는 IoT 서비스로 전달된다. 반대로 클라우드에 있는 IoT 서비스는 단말(스포츠 시계 또는 스마트폰)을 연결하고 관리한다. 클라우드로 전달된 데이터는 저장되어 데이터 분석서비스에서 활용된다. 클라우드가 제공하는 IoT 서비스의 핵심은 기기와 연결 및 관리 기능과 데이터 수집 및 분석 기능이라고 보면 된다. 특히 클라우드는 쉽게 글로벌하게 확대가 가능하기 때문에 해당 기기가 전세계적으로 서비스되더라도 쉽게 IoT 서비스를 제공할 수 있다.

〈그림 11-52〉 클라우드의 IoT 서비스 동작



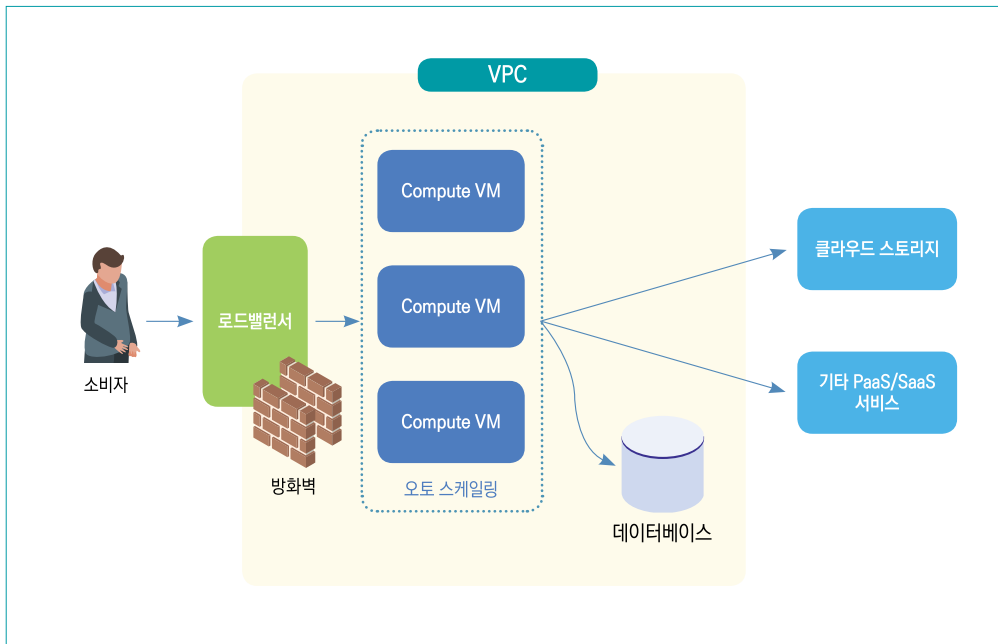


## 9 네트워크/보안

네트워크와 보안은 전문적인 IT 기술이 필요한 영역이라 약간은 어렵게 느껴질 수 있지만 핀테크를 위한 IT 시스템 구성에 필수적인 부분이라 지나칠 수 없는 부분이다. 클라우드에서는 이러한 전문적인 지식이 필요한 네트워크와 보안을 쉽게 구성할 수 있도록 했으며, 특히 기본적인 구성이 어느 정도 되어 있어 쉽게 시작할 수 있도록 되어 있다. 네트워크 서비스의 주요 종류는 VPC(Virtual Private Cloud), 로드밸런서 등이 있다. 보안 서비스는 주로 방화벽(firewall), 침입 탐지(intrusion detection), 계정 관리 등이 있다.

먼저 VPC는 컴퓨팅 자원이나 데이터베이스 등이 서로 통신을 하기 위해 사용하는 기본적인 네트워크를 말한다. 컴퓨팅 자원인 VM을 사용하려면 VPC가 반드시 있어야 하며, 클라우드에서는 보통 기본적으로 VPC를 만들어서 제공하므로 별도의 구성은 필요하지 않다. 만약 추가적인 구성이 필요하면 IT 전문가와 상의하여 진행하는 것이 좋다. 동일한 VPC 안에 있는 IT 자원들은 회사나 가정 내에 있는 IT 기기들처럼 특별한 작업을 하지 않아도 기본적으로 통신이 된다고 보면 된다. 로드밸런서 서비스를 통해 사용자들의 부하가 대량으로 몰려도 클라우드상에 구현된 컴퓨팅 자원들이 부하를 나눠 처리하고 필요시 자동 확장(auto scaling)을 하게 한다. 그리고 방화벽을 통해 외부에서의 침입을 차단한다. 침입 탐지 기능은 외부에서 해킹 등을 통해 뚫고 들어온 침입자를 탐지하여 관리자에게 알리는 기능을 의미한다. 핀테크는 금융서비스를 근간으로 하고 있기 때문에 외부 침입으로 인한 피해가 치명적일 수 있다. 따라서 이러한 서비스들은 눈에 들어나지는 않지만 필수적이므로 반드시 전문가와 상의를 통해 적절하게 구축이 되어 있어야 한다.

〈그림 11-53〉 일반적인 클라우드의 네트워크 및 보안 아키텍처




## 1 클라우드 서비스 가입

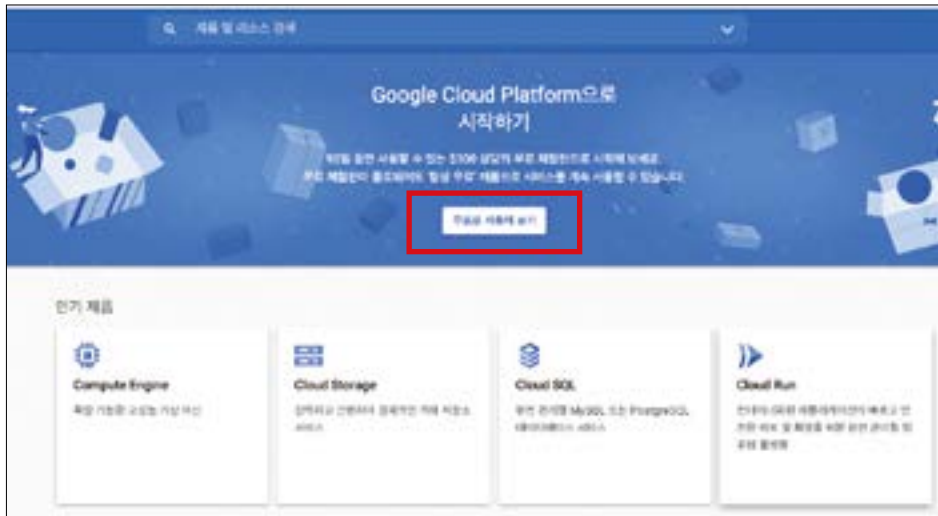
클라우드의 시작은 온라인 서비스 가입부터 시작한다. 구글 GCP는 아래 웹사이트에서 회원 가입을 할 수 있다.

<https://console.cloud.google.com/>

URL을 입력하고 들어가면 먼저 아래와 같이 로그인을 하라는 페이지가 뜬다. 만약 이미 로그인된 상태라면 이 과정은 넘어가도 된다. 만약 계정이 없다면 &계정 만들기'를 눌러 계정을 만들어야 한다.



로그인을 하고 나면 다음과 같은 페이지가 보인다. 구글 GCP를 최초로 가입하면 \$300의 무료 크레딧을 제공한다. 그러기 위해서는 결제 계정에 신용카드 정보를 입력해야 하는데, 그렇다고 해서 비용이 청구되는 것은 아니고 실제 유료 전환은 반드시 사용자의 동의 과정을 거쳐야 하므로 안심해도 된다. '무료로 사용해 보기'를 클릭한다.



약관에 동의하고 개인정보 인증 과정을 거치고 나면 결제 정보 입력하는 부분이 나온다. 만약 구글 Play에서 이미 결제를 하고 있는 부분이 있다면 그것을 이용해도 되고, 새로운 프로필을 만들 수도 있다. 여기서는 새로운 프로필을 만든다.

결제 프로필 설정에서 사업체의 경우에는 계정 유형을 사업자로, 개인의 경우에는 개인을 선택해서 진행하면 된다. 차이가 있다면 사업자인 경우에는 사업자 번호를 놓고 신용카드는 법인카드를 사용해야 한다는 점에 주의하자. 개인의 정보를 꼼꼼하게 입력하고 계속을 눌러 다음으로 넘어간다.

**결제 프로파일** ⓘ  
이 계정이나 계정에 연결된 결제 프로파일 (결제카드) 또는 Google 계정에 대해 프로젝트에 공유된 신용카드는 없습니다.

**결제 프로파일 만들기**

**계정 정보**

**계정 유형** ⓘ ⓘ / ✎  
개인

**세금 정보** ⓘ  
세금사 유형: 개인

**연락처 주소** ⓘ

회사

도시

주소

주소 입력란 1

주소 입력란 2

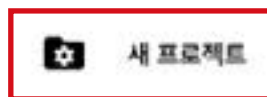
이름  
**홍길동**

휴대번호 ⓘ ⓘ

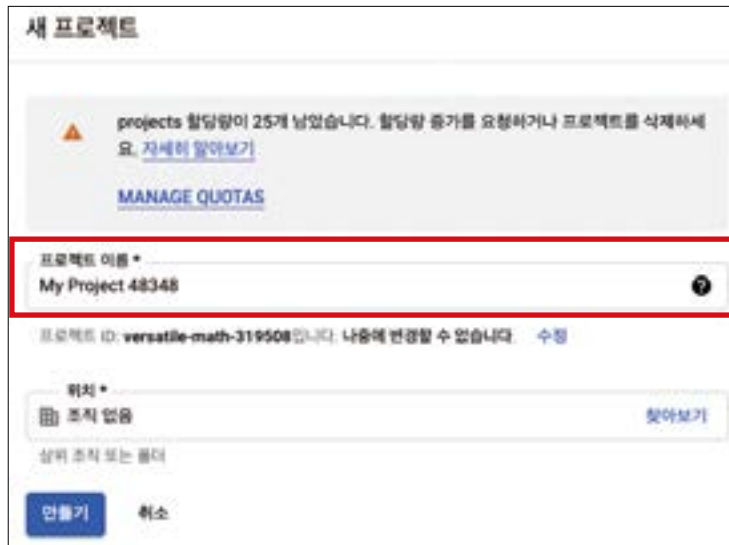
전화번호 (선택사항)

그러면 개인정보가 맞는지 통신사 인증을 통해 확인한다. 그리고 나서 신용카드 정보를 넣고 결제(신용카드) 정보를 완성하면 가입이 완료된다. 가입이 되었으면 구글 GCP를 사용하기 위한 첫 번째 프로젝트를 생성한다. 프로젝트는 클라우드상의 리소스를 관리하는 단위로 기본적으로 하나가 필요하다. 프로젝트 생성은 클라우드 콘솔 화면에서 상단의 이 부분에서 눌러 시작한다.

창이 뜨면 우측 상단의 새 프로젝트를 클릭한다.

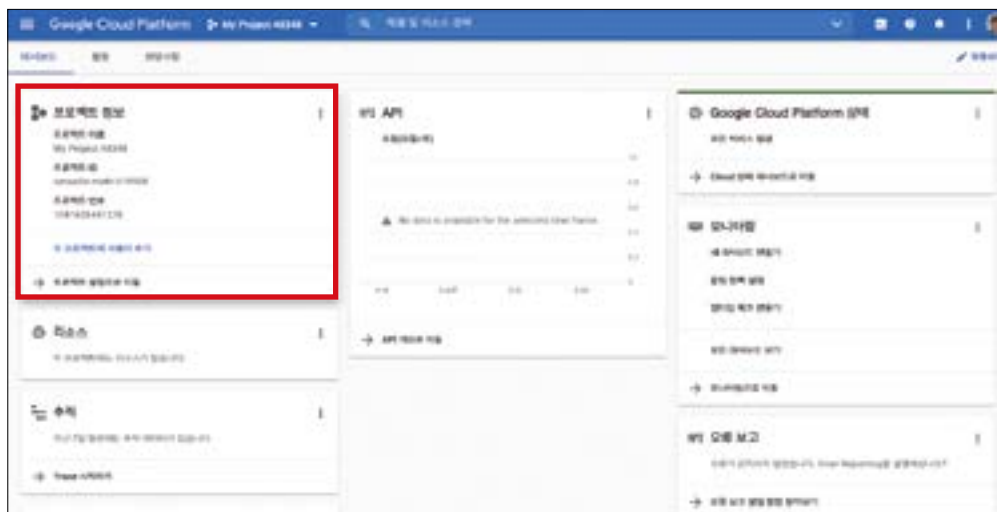


프로젝트 이름에는 원하는 이름을 입력한다. 그리고 만들기를 클릭한다.



약간의 시간이 걸려 프로젝트가 생성되면 다시 상단의 이 부분을 클릭하여 방금 만들어진 프로젝트를 선택한다.

다음과 같이 프로젝트 대시보드가 표시되면 이제 구글 GCP 프로젝트를 사용할 준비가 된 것이다.



## 2 컴퓨팅 자원 생성

다음은 실제로 구글 GCP 프로젝트에 필요한 자원을 만들어 보자. 여기서는 컴퓨팅 VM을 생성하고 웹서버를 하나 설치해서 실제로 브라우저에서 불러보는 과정을 진행할 것이다. 구글 GCP에서 원하는 서비스 메뉴는 아래에 보이는 햄버거 메뉴를 눌러 선택할 수 있다. 컴퓨팅 자원은 Compute Engine이다.

만약 서비스를 잘 찾지 못하겠으면 검색창에 입력해서 원하는 메뉴를 클릭해도 된다.



인스턴스 만들기를 누른다.



기본적인 정보를 아래와 같이 선택한다. 서울 리전에 가상 CPU 2개, 1GB 메모리를 가진 VM을 my-first-vm이라는 이름으로 생성하는 것이다.



하단의 방화벽 부분에서 HTTP, HTTPS 트래픽을 허용하는 것으로 체크한다. 나머지는 디폴트 설정 그대로 만들기를 누른다.



다음과 같이 my-first-vm이 뜨면 상태가 녹색으로 변한다. 그러면 VM 접속을 위해 SSH를 클릭한다.



다음과 같이 브라우저의 새로운 탭이 만들어지며 텍스트 터미널 프롬프트가 보이면 SSH 접속에 성공한 것이다. SSH는 VM에 접속해서 필요한 명령을 수행할 수 있는 기능이다.



다음 명령을 입력하여 웹서버를 설치한다.

```
sudo apt-get install -y apache2
```



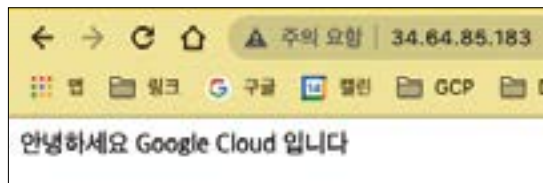
다음과 같이 웹서버를 테스트하기 위한 HTML을 입력한다.

```
cat > /var/www/html/index.html <<EOF
<html>
<meta charset='UTF-8'>
안녕하세요 Google Cloud 입니다
</html>
EOF
```

다시 브라우저에서 클라우드 콘솔 탭으로 이동한다. VM의 외부 IP 부분을 클릭한다.

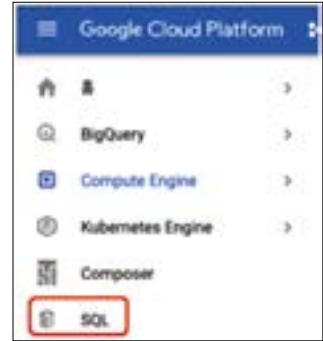


아래와 같이 내용이 브라우저에 보이면 성공이다. 위에 입력한 HTML의 내용이 VM에 설치된 웹서버를 통해 서비스되고 그것을 브라우저에서 접속해서 출력한 것이다. 여기서는 간단한 HTML 페이지를 보이는 작업만 진행했지만 원하는 프로그램이나 소프트웨어를 추가로 설치하면 보다 복잡한 서비스를 구현할 수 있다.



### 3 데이터베이스 서비스

데이터베이스 서비스를 하나 만들어 보자.  
데이터베이스는 아래와 같이 SQL 메뉴를 선택한다.



인스턴스 만들기를 클릭한다.



MySQL을 선택한다.



다음과 같이 정보를 입력하고 데이터베이스를 생성한다. 실제 생성까지는 수 분 정보가 소요되므로 커피 한 잔 하면서 기다리는 것도 좋다.

### 인스턴스 정보

인스턴스 ID \*

my-first-db

스모크, 손지, 워터마크를 사용해서도 문자로 시작해지 않습니다

비밀번호 \*

\*\*\*\*\* 👁️ 생성

모든 사용자의 내일번호를 설정하십시오. [자세히 알아보기](#)

비밀번호 없음

데이터베이스 버전 \*

MySQL 5.7

### 리전 및 영역 가용성 선택

성능을 향상시키려면 필요한 서비스의 가까운 위치에 데이터를 보관하십시오. 리전은 영구적이지 않은 영역은 변경할 수 없습니다.

리전

asia-northeast3(서울)

단일 영역  
서비스 중단에 영향을 경우 장애 주기가 짧습니다. 프로비전에는 권장되지 않습니다.

여러 영역(고가용성)  
선역한 리전 내의 다른 영역으로 자동으로 장애 호시가 적용됩니다. 프로비전 인스턴스에 권장되며 비용이 증가합니다. 바이너리 로그(백업에 필요) 및 자동 백업을 사용 설정합니다. 7일 이상의 로그를 지원할 수 있는 저장 용량이 있거나 없습니다.

▼ 영역 시열

### 인스턴스 맞춤설정

다음에 인스턴스 구성을 맞춤설정할 수도 있습니다.

▼ 구성 옵션 표시

인스턴스 만들기
취소

아래와 같이 상태가 녹색으로 바뀌면 생성이 완료된 것이다.



왼쪽의 메뉴에서 데이터베이스를 선택한다. 여기서 데이터베이스는 데이터베이스 서비스 자체를 말하는 것이 아니라 MySQL 데이터베이스 내에서 여러 개의 테이블들을 묶어서 관리하는 단위를 말한다. 최초 생성 시에는 시스템에서 만든 디폴트 데이터베이스밖에 없으므로 업무용으로 사용할 데이터베이스를 하나 생성한다.



목록이 보이면 데이터베이스 만들기를 클릭한다.

이름 ↑	대조	문자 집합	유형	
information_schema	utf8_general_ci	utf8	시스템	⋮
mysql	utf8_general_ci	utf8	시스템	⋮
performance_schema	utf8_general_ci	utf8	시스템	⋮
sys	utf8_general_ci	utf8	시스템	⋮

다음과 같이 입력하고 만들기를 누른다. 그러면 목록에 새로운 데이터베이스, mydb가 추가된 것이 보일 것이다.

### 데이터베이스 만들기

데이터베이스 이름 \*

MySQL 식별자 규칙을 따라야 합니다. [자세히 알아보기](#)

문자 집합 \*

나중에 ALTER DATABASE 쿼리를 실행하여 변경할 수 있습니다.

대조

테이블을 생성하기 위해 my-first-db 데이터베이스에 접속한다. 왼쪽의 메뉴에서 개요를 눌러 들어간 후에 아래와 같은 화면에서 CLOUD SHELL 열기를 눌러 클라우드 셸을 실행한다.



클라우드 셸에 들어가면 아래 명령이 실행 대기 중으로 보이는데 엔터키를 눌러 실행한다. 권한을 물어보는 창이 뜨면 승인을 눌러 허용한다.

```
gcloud sql connect my-first-db --user=root --quiet
```

1~2분 정도의 시간이 지나면 다음과 같이 암호를 입력하라는 프롬프트가 뜬다. 데이터베이스 생성 과정에 입력했던 암호를 여기서 입력한다. 정상적으로 암호가 입력되었으면 mysql> 이라는 프롬프트가 뜬다. 이제 SQL을 실행할 수 있는 준비가 된 것이다.

```
Connecting to database with SQL user [root].Enter password: *****
mysql>
```

아래와 같은 데이터를 입력할 테이블을 만들어 볼 것이다.

	A	B	C	D	E	F
1	Date	Item	Count	UnitPrice	TotalCost	User
2	2021-01-01	컴퓨터	10	1,000,000	10,000,000	홍길동
3	2021-01-02	컴퓨터	1	1,000,000	1,000,000	이순신
4	2021-01-03	모니터	5	200,000	1,000,000	홍길동
5	2021-01-04	모니터	2	200,000	400,000	신사임당
6	2021-01-05	키보드	20	5,000	100,000	홍길동
7	2021-01-06	키보드	2	5,000	10,000	신사임당
8	2021-01-07	책상	1	500,000	500,000	이순신
9	2021-01-08	컴퓨터	3	1,000,000	3,000,000	홍길동
10						
11						
12						

테이블을 생성하려면 mysql> 프롬프트에서 다음 내용을 입력한다.

```
use mydb;  
create table transaction (  
  TrDate date,  
  Item varchar(50),  
  Count int,  
  UnitPrice decimal,  
  TotalCost decimal,  
  User varchar(20)  
);
```

CSV 파일에서 데이터를 로딩한다. 원래는 각자 원하는 데이터를 로딩하는게 맞겠지만 여기서서는 테스트를 위해 미리 제작해 둔 CSV 파일을 로딩한다. 콘솔의 왼쪽 메뉴에서 개요를 누른다. 그리고 가져오기를 누른다.



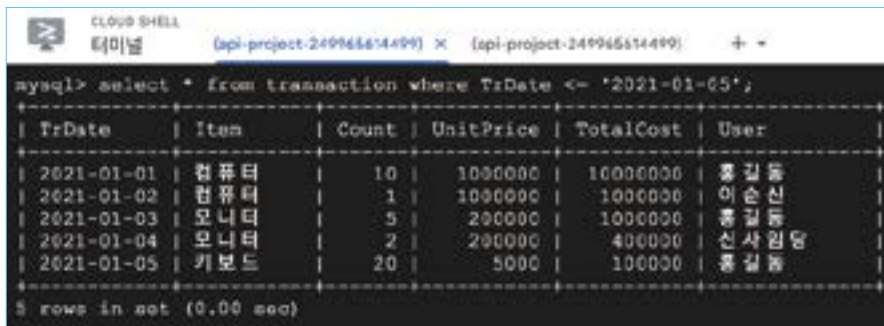
bucket-name/file-name 부분에 gs://javalove93-samples/table\_transaction/transaction.csv를 다음과 같이 입력하고 가져오기를 누른다. 10~20초 정도 지나면 데이터 로딩이 완료된다.



그러면 하단의 클라우드 셸 mysql> 프롬프트에서 다음과 같이 2021년 1월 5일 이전의 데이터를 조회해 본다.

```
select * from transaction where TrDate <= '2021-01-05';
```

그러면 다음과 같이 조회된다. SQL 문의 WHERE 조건에 해당하는 부분을 바꿔가며 다른 데이터를 조회해 보기 바란다.



TrDate	Item	Count	UnitPrice	TotalCost	User
2021-01-01	컴퓨터	10	1000000	10000000	홍길동
2021-01-02	컴퓨터	1	1000000	1000000	이순신
2021-01-03	모니터	5	200000	1000000	홍길동
2021-01-04	모니터	2	200000	400000	신사임당
2021-01-05	키보드	20	5000	100000	홍길동

## 4 서버리스 코드 서비스

클라우드에서는 애플리케이션을 올리기 위한 방법을 여러 가지 제공한다. 첫 번째 실습에서 본 컴퓨팅 VM이 가장 기본적인 서비스이다. VM의 장점은 클라우드 이전의 시스템에서 사용하던 하드웨어 서버와 거의 동일하여 익숙하다는 것이다. 반면 클라우드의 장점인 확장성, 민첩성 면에서는 제약사항이 존재한다. 그런 면에서 서버리스 코드는 사용자가 프로그램만 입력하면 해당 애플리케이션에 대한 복잡한 관리, 설치 등이 없이도 고도의 확장성과 안전성을 제공하는 장점이 있다. 특히 애플리케이션을 작성하기 위해 필요한 프레임워크(Apache 웹서버)들을 설치할 필요 없이, 프로그램 코드만을 올리면 바로 서비스가 가능하기 때문에 IT 엔지니어가 많지 않은 경우에도 쉽게 서비스를 올릴 수 있는 장점이 있다.



여기서는 아마존 AWS가 제공하는 서버리스 서비스인 Lambda에 대해 실습을 해보려고 한다. AWS는 구글 GCP와 마찬가지로 무료 서비스를 제공하는데 방법은 약간 다르다. GCP가 일정 비용을 무료로 제공하는 데 반해, AWS는 계정 생성 후 낮은 사양의 리소스에 대해서는 무료로 제공하는 정책을 사용한다. Lambda의 경우에는 서비스 호출 회수로 과금을 하는데, 월 1백만 회까지는 무료로 서비스를 사용할 수 있다. 그러면 AWS 계정 생성 방법부터 알아보겠다.

AWS 계정 생성은 여기에서부터 시작한다.

```
https://aws.amazon.com/
```

우측 상단의 AWS 계정 생성을 클릭한다.

AWS 계정 생성

그리고 이메일 주소, 암호, 주소 등의 정보를 입력하면 된다. 사용자 종류는 개인과 비즈니스 중에서 특별한 이유가 없다면 개인으로 선택하고 신용카드 정보를 입력한다. 마지막으로 SMS 인증을 하면 가입절차는 완료된다. 그리고 나서 AWS 콘솔로 이동한다.

```
https://aws.amazon.com/console
```

그리고 나서 우측 상단에 있는 리전(지역)을 눌러 Seoul 리전을 선택한다. 그러면 이 실습에서 만드는 리소스들은 서울 리전에 생성된다.

Ohio ▼Asia Pacific (Seoul) ap-northeast-2

Lambda 서비스는 다음과 같이 검색 창에 'lambda' 입력 후 검색 결과의 Lambda를 선택하여 이동한다.



다음은 Lambda 서비스 콘솔이다. 여기서 만드는 서버리스 코드를 함수(function)이라고 하는데, 아직은 함수가 만들어지기 전이다. Create Function을 눌러 함수를 생성한다.



Author from scratch를 선택한다.



실행할 함수의 기본적인 정보를 다음과 같이 입력한다. 사용할 프로그래밍 언어로는 Python 3.8을 선택하며, 만약 다른 언어를 사용하고 싶다면 그것을 선택하면 된다. 그리고 나서 하단의 Create function 버튼을 누른다.



The screenshot shows the 'Basic information' section of the AWS Lambda console. It includes the following fields:

- Function name:** A text input field containing 'helloworld'.
- Runtime:** A dropdown menu set to 'Python 3.8'.
- Permissions:** A section with a link to 'Change default execution role'.

그러면 다음과 같이 helloworld 함수가 만들어진다. 여기서 Add trigger를 눌러 웹 브라우저를 통한 서비스가 가능하도록 설정한다.



아래와 같이 API Gateway를 선택하고, API 종류로는 HTTP API, Security는 Open(인증 없이 실행 가능)을 선택하면 된다.

**Add trigger**

**Trigger configuration**

**API Gateway**  
api application-services aws serverless

Add an API to your Lambda function to create an HTTP endpoint that invokes your function. API Gateway supports two types of RESTful APIs: HTTP APIs and REST APIs. [Learn more](#)

**API**  
Create a new API or attach an existing one.

Create an API

**API type**

**HTTP API**  
Create an HTTP API.

**REST API**  
Create a REST API.

**Security**  
Configure the security mechanism for your API endpoint.

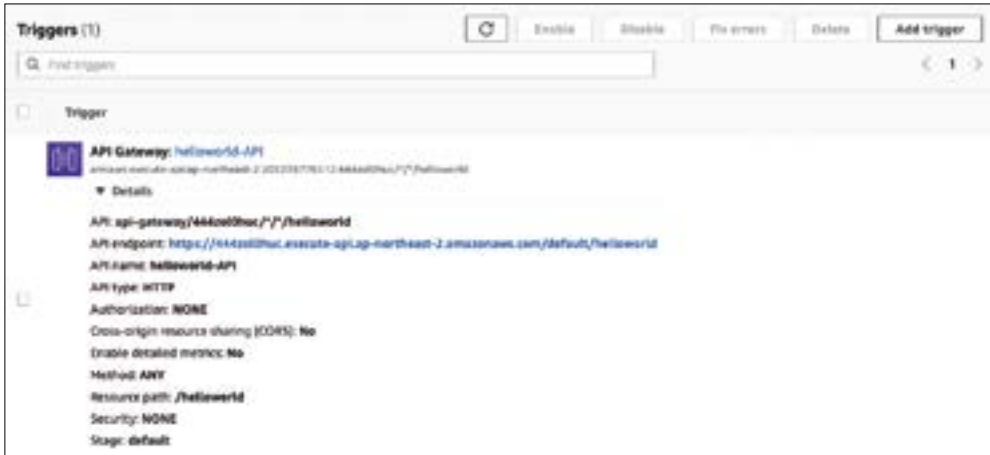
Open

▶ **Additional settings**

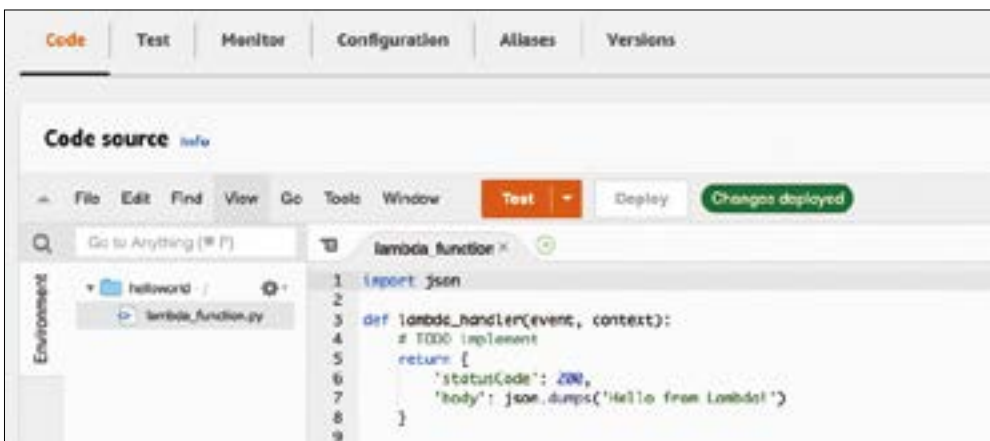
Lambda will add the necessary permissions for Amazon API Gateway to invoke your Lambda function from this trigger. [Learn more](#) about the Lambda permissions model.

Cancel **Add**

그러면 하단에 아래와 같이 웹브라우저로 실행가능한 정보들이 보이며, 이 중에서 API endpoint에 해당하는 URL을 복사하여 브라우저 주소창에 입력하고 실행하면 된다.



현재는 함수가 만들어질 때, 기본적으로 만들어지는 코드가 프로그램으로 실행이 되며, 그것을 각자 원하는 코드로 수정을 해 본다. Code 탭을 눌러서, 왼쪽의 탐색기에서 lambda\_function.py를 클릭하면 메인 창에 기본적인 샘플 코드가 보인다.



그것을 다음과 같이 수정한다. 여기서는 간략하게 영문으로 되어 있는 메시지를 한글로 수정하기만 한다. 실제로는 각자의 비즈니스 로직을 넣어야 한다.

```
1 import json
2
3 def lambda_handler(event, context):
4     # TODO implement
5     return {
6         'statusCode': 200,
7         'body': '안녕하세요 람다 서비스입니다'
8     }
9
```

그리고 나서 변경된 코드를 반영하기 위해 Deploy 버튼을 누른다.



API gateway 항목에서 보았던 URL 주소를 이용하여 브라우저로 페이지를 열어 본다. 아래와 같이 한글 메시지가 보이면 성공한 것이다.



이렇게 만들어진 AWS Lambda 서비스는 간단해 보이지만, 동시에 100만명이 접속해도 서비스가 가능할 만큼 강력한 확장성과 안전성을 제공한다. IT 개발자와 상의하여 필요한 비즈니스 로직을 추가하기만 하면 바로 여러분들이 원하는 서비스를 제공할 수 있게 된다.

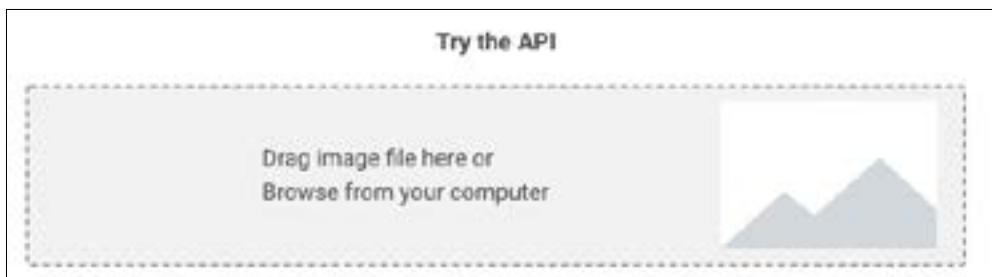
## 5 AI 서비스

지금까지 클라우드 서비스들을 실습해 보았고, 이번에는 인공지능 AI/ML(Artificial Intelligence, Machine Learning) 기반의 재미 있는 서비스를 하나 실습해 볼까 한다. 사실 AI/ML 기능은 전문가가 데이터를 가지고 학습을 하여 모델을 만들어야 하고, 이 과정은 수 일에서 수 주 혹은 그 이상 걸릴 수 있는 작업이다. 하지만 클라우드에서는 이러한 모델 개발 과정을 거치지 않고 바로 서비스에 적용할 수 있는 AI/ML API를 제공한다. 여기서는 그중에서 이미지 인식 기능에 대해 실습해 보겠다. 실제로는 이 서비스를 이용할 수 있게 프로그램을 작성해야 하지만, 그 부분은 다소 복잡할 수 있으므로 여기서는 브라우저상에서 바로 테스트해 볼 수 있는 데모를 이용할 것이다. 이러한 데모를 손쉽게 제공하는 구글 GCP를 이번 실습에 사용한다.

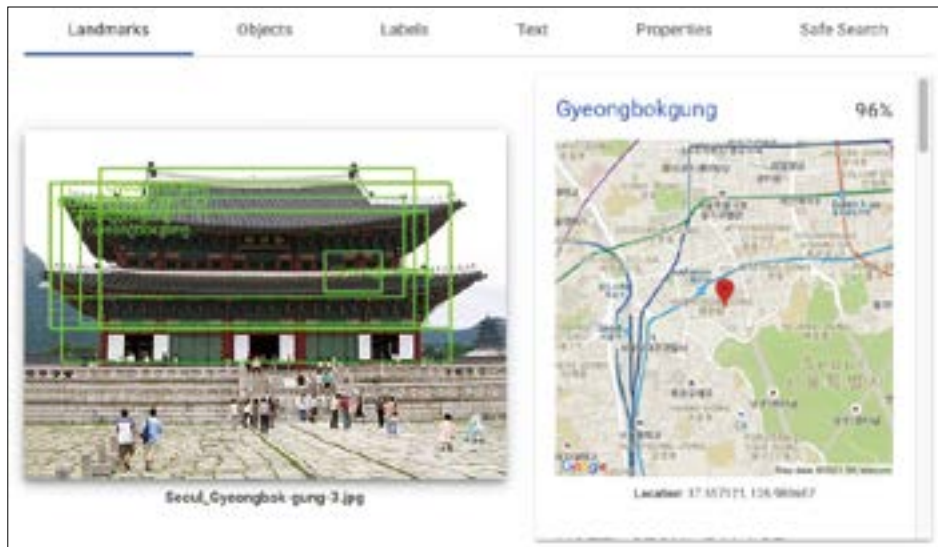
먼저 이미지 인식을 위해서 아래 URL을 브라우저에 입력한다.

<https://cloud.google.com/vision>

브라우저 중간 쪽에 아래와 같이 바로 테스트를 해 볼 수 있는 기능이 보일 것이다. 여기에 아무 이미지나 가져다가 넣어 볼 수 있는데, 탐색기에서 파일을 드래그해서 넣을 수도 있고, 또는 네모 박스 부분을 클릭해서 자신의 컴퓨터에 있는 이미지를 찾아 업로드할 수 있다. 어느 방법이든 이미지를 하나 골라 입력해 보기 바란다.



필자는 경복궁 이미지를 하나 골라 입력해 보았다. 그러면 이미지의 다양한 정보를 판독해서 제공하는데, 첫 번째는 경복궁 자체가 대중들에게 많이 알려져 있는 건물(랜드마크)이므로 그것을 먼저 알려준다. 지도의 위치(좌표)까지 표시된다. Objects 탭을 선택하면 이 사진에 포함된 사물들이 어떤 것들이 있는지 보여준다. Text는 경복궁 현판(육안으로 잘 안보일 수 있지만 사진에 한문으로 되어 있는 글씨가 포함되어 있음)의 글씨를 OCR로 판독을 시도하여 결과를 보여준다. Safe Search는 혹시라도 이 그림에 폭력, 성인, 인종 등에 관련된 좋지 않은 내용이 있는지 판독해 준다.



이처럼 AI/ML 모델 개발 작업 없이도 이미지로부터 많은 정보를 판독할 수 있다. 다만, 이러한 모델은 일반적인 이미지를 학습한 결과이므로 특정 인물에 대한 얼굴인식(face recognition)과 같은 기능은 불가능하다. 그런 경우는 앞에서 언급한 모델 개발 작업이 필요하다.



이 외에도 음성 인식(STT, Speech-To-Text), 음성 출력(TTS, Text-To-Speech), 자연어 분석 등의 기능도 역시 API로 제공하므로 데모를 한 번 해 보기 바란다.

STT: <https://cloud.google.com/speech-to-text>

TTS: <https://cloud.google.com/text-to-speech>

자연어: <https://cloud.google.com/natural-language>

## 6 서비스 비교표

클라우드의 서비스들에 대해 직접 실습하는 기회를 가져 봤는데, 지면의 제약상 특정 CSP 서비스들만을 사용해 보았다. 실제로는 CSP들이 제공하는 서비스들이 이름은 다르지만 동일하거나 유사한 서비스를 제공하는 경우가 많다. 이러한 서비스들과 각 CSP가 제공하는 이름은 무엇인지 살펴보도록 하겠다.

〈표 II-8〉 클라우드 서비스 비교표

기능	AWS	Azure	GCP
컴퓨팅 VM	EC2	AVM	GCE
스토리지	S3	Blob Storage	GCS
데이터베이스	RDS	Azure Database	Cloud SQL
빅데이터 분석	RedShift	Synapse	BigQuery
서버리스 코드	Lambda	Azure Functions	Cloud Function
인공지능 AI/ML	SageMaker	Azure ML	Vertex AI
MSA/Kubernetes	EKS	AKS	GKE
IoT	IoT Core	IoT Hub	Cloud IoT
메시징	SMS	Service Bus Messaging	Pubsub

헬로, 핀테크!(핀테크 기반기술)



HELLO, FINTECH!

# 3장

## IoT

제1절 처음 만나는 IoT

제2절 IoT의 탄생

제3절 IoT의 기능과 역할, 구조, 주요 특징

제4절 IoT 설계

제5절 IoT와 핀테크

제6절 IoT 미래 전망

제7절 IoT와 좀 더 친해지기

# 3장 IoT



## 💡 학습목표

- ① IoT 센서 기술의 구조와 특징을 열거할 수 있다.
- ② IoT 통신기술의 구조와 특징을 열거할 수 있다.
- ③ IoT 기술의 중요성과 향후 기술발전 사항에 대해 설명할 수 있다.
- ④ IoT의 보안적 이슈와 이를 해결하기 위한 요소 기술에 대해 설명할 수 있다.
- ⑤ IoT 기술과 서비스 관련 표준화 동향에 대해 설명할 수 있다.
- ⑥ 산업용 IoT(IIoT)의 정의와 이에 대한 기술적 특징에 대해 설명할 수 있다.
- ⑦ 핀테크에 활용 중이거나 향후 적용될 수 있는 IoT 기술에 대해 설명할 수 있다.
- ⑧ IoT 기술을 핀테크 비즈니스 실무에 적용할 수 있다.

## 💡 학습개요

4차 산업혁명과 함께 제조, 건설, 에너지 등의 전통적인 산업 뿐만아니라 금융, 보험 등에서도 디지털화가 빠르게 진행되고 있다. 기존 비즈니스를 한단계 업그레이드하여 매출을 증대시키거나 비용효율을 높이기 위해, 또는 신규 비즈니스를 창출하기 위해 디지털트랜스포메이션은 많은 산업과 기업에게 당면과제이다. 이를 가능하게 한 것 중 중요 요소기술이 바로 사물인터넷(IoT) 기술이다. IoT 기술이 대두될 수 있었던 것은 빠르게 발전하는 센서, 통신, 클라우드, 데이터 분석 등의 기술에 의해서이다.

이 장에서는 IoT의 정의와 개념에 대해 살펴보고 등장 배경과 발전 역사에 대해 알아보겠다. 또한 센서, 연결(통신), 수집(클라우드), 분석(빅데이터), 실행(앱 등)으로 이뤄진 IoT 프레임워크에 대해서도 알아보고, 특히 센서와 연결 계층에 대한 기술과 특징에 대해 설명하겠다. 또한, IoT 기술을 설계하고 산업에 적용하기 위하여 도입 사례, 보안상 고려사항, 표준화 동향 등도 알아보고자 한다. 이러한 IoT 기술이 현재와 향후 핀테크 산업에 어떻게 적용할 수 있는지 실제 비즈니스 사례와 향후 핀테크 혁신 방안을 토대로 설명하겠다. 끝으로 IoT의 이해를 돕기 위해 NFC 센서를 이용한 IoT 솔루션 개발 내용을 설명하고, 최신 IoT 트렌드 중 하나인 AR, VR, 메타버스 서비스에 대해 알아보겠다.

 용어해설

① 5G

5G는 5세대 무선 네트워크 기술. 4G 이후의 차세대 글로벌 무선 표준인 5G는 더 높은 데이터 속도, 더 짧은 지연시간 및 더 많은 사용자, 디바이스, 서비스를 지원하는 동시에 네트워크 효율성을 향상시키도록 설계되었음

② MEMS(Micro Electro Mechanical System)

‘초소형 전기기계 시스템’이라는 뜻. 실리콘 기판이나 유리 기판, 또는 유기소재 등의 위에 기계요소 부품으로서 센서, 액츄에이터, 전기회로 등을 한 자리에 탑재한 디바이스를 가리킴(총 길이가 mm단위, 그 속의 부품은  $\mu\text{m}$  단위)

③ SoC(System on Chip)

하나의 집적회로에 집적된 컴퓨터나 전자 시스템 부품을 가리킴. 디지털 신호, 아날로그 신호, 음성 신호와 RF 기능 등이 단일 칩에 구현되어 있음. 일반적으로 임베디드 시스템 영역에 주로 사용됨

④ MCU(Micro Controller Unit)

MCU는 마이크로프로세서와 입출력 모듈을 하나의 칩으로 만들어 정해진 기능을 수행하는 컴퓨터를 말함. CPU 코어, 메모리 그리고 프로그램 가능한 입/출력을 가지고 있음

⑤ NFC(Near Field Communication)

근거리 무선 통신을 뜻하는 기술로, 두 대 이상의 단말기를 10cm 이내로 접근시켜 양방향 데이터를 송수신하는 기술

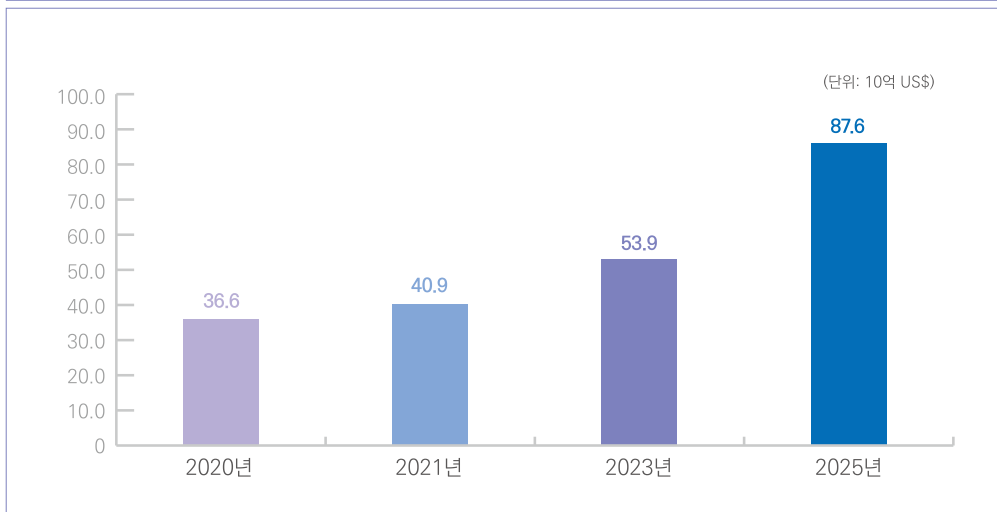
⑥ LPWA(Low Power Wide Area)

저전력 장거리 통신기술. 통신반경이 수십 km로 넓고 전력소모가 적어 단말기의 배터리 수명이 수년간 유지되는 장점을 가지고 있어 온도, 습도, 무게, 위치 등 단순 정보를 측정해 처리하는 데 유용하게 활용될 수 있음

## 1 IoT(사물인터넷)의 정의와 개념

사물인터넷(IoT; Internet of things)은 가전제품, 건설장비, 자동차, 웨어러블 디바이스 등 각종 사물에 센서와 통신 기능을 내장하여 인터넷에 연결하는 기술이다. 또한 무선통신, 인터넷 등으로 연결된 사물들이 전송하는 데이터를 분석하고 학습한 정보를 사용자에게 제공하거나 사용자가 이를 바탕으로 원격 조정할 수 있는 기술 또한 포함한다. 사물인터넷에 연결되는 사물들은 자신을 구별할 수 있는 유일한 아이피(IP)를 가지고 인터넷으로 연결된다.

〈그림 III-1〉 세계 IoT 센서 시장 규모 및 전망



출처: MarketandMarkets(2020. 5.)

글로벌 리서치회사 마켓츠앤마켓츠(MarketsandMarkets)에 따르면 IoT 센서 시장 규모는 2020년 366억 US 달러에서 2025년 876억 US 달러로 커질 것으로 예상되며, 연평균 19.0%의 성장률을 보일 것으로 전망했다. 스마트시티, 스마트홈, 자율 주행과 5G 등 사물과 사물이 직접 상호작용하는 IoT 관련 부품 수요 증가세에 맞추어 IoT 센서의 수요도 크게 늘어날 것으로 전망된다. 이와 같이 많은 사물이 연결되면 인터넷을 통해 방대한 데이터가 모이게 되는데, 이렇게 모인 데이터는 기존 기술로 분석하기 힘들 정도로 방대해진다. 이러한 방대한 데이터, 즉, 빅데이터를 분석하는 효율적인 알고리즘을 개발하는 기술의 필요성이 사물인터넷의 등장에 따라 함께 대두되고 있다.

## 2 IoT vs. IoP

기존의 인터넷은 근본적으로 사람과 사람을 연결하는 Internet of People(IoP)의 형태였다. 하지만 사물과 인간은 다르기에 IoT와 IoP 또한 다음과 같은 근본적인 세 가지 차이점이 존재한다.

### #1. 사물은 인간보다 월등히 많다

IT 관련 신문에서 머지않은 미래에 수많은 사물이 인터넷에 연결될 것이라는 기사를 많이 보았을 것이다. 미국 시스코社の CEO이었던 존 챔버스(John Chambers)는 2024년까지 사물 5,000억 개가 연결될 것이라고 예측했다. 이 측정 수치는 전 세계 인구의 약 100배이다.

### #2. 사물은 인간보다 훨씬 더 많은 것을 말해 준다

사람이 데스크탑, 휴대전화 등 다양한 디바이스를 활용해 애플리케이션을 사용할 때 주요 데이터 Input 매개체는 키보드이다. 그리고 애플리케이션은 대부분 어떤 형태를 사용하여 사람으로부터 간단한 데이터를 수집한다. 그에 비해 사물은 센서가 훨씬 더 많다. 일반적인 휴대전화의 센서는 가속도계, GPS, 심지어 방사선 검출기까지 포함해 대략 14개이다. 풍력 터빈, 유전자 시퀀서, 고속 삽입기 등과 같은 산업용 사물의 센서는 대부분 100개 이상이다.

### #3. 사물은 멈추지 않는다

인간 인터넷(loP; Internet of People) 애플리케이션의 데이터 대부분은 인간이 뭔가를 구매하거나, 업무하는 등에 관련된 프로세스 중 일부에 해당한다. 그러나 인간은 전자 상거래, 인사 관리(HR), 구매, 고객 관계 관리(CRM; Customer Relationship Management), 기업 자원 관리(ERP; Enterprise Resource Planning) 애플리케이션에 데이터를 자주 입력하지 않는다. 반면에 전력망 센서는 초당 60회, 건설 지게차는 분당 1회, 고속 인서터는 2초마다 한 번씩 데이터를 전송할 수 있다.

이러한 loP와의 세 가지 차이점 때문에 IoT 기술의 파급력은 상당할 것으로 기대된다. 엄청나게 많은 수의 사물 센서들이 끊임없이 많은 빅데이터를 생성함에 따라, 이를 이용한 새로운 비즈니스와 제품 및 서비스가 탄생하고 있다.



## 1 IoT의 등장 배경

제조, 건설, 에너지 등의 전통적인 산업뿐만 아니라 금융, 보험 등에서도 디지털화가 빠르게 진행되고 있다. 기존 비즈니스를 한 단계 업그레이드하여 매출을 증대시키거나 비용효율을 높이기 위해, 또는 신규 비즈니스를 창출하기 위해 디지털트랜스포메이션은 많은 산업과 기업에게 주어진 당면과제이다. 이를 가능하게 한 것 중 중요 요소기술이 바로 IoT 기술이다. IoT 기술이 대두될 수 있었던 것은 빠르게 발전하는 센서, 통신, 클라우드, 데이터 분석 등의 기술에 의해서이다.

미래 최고의 기업은 다양한 기기가 연결된 새로운 세상을 완벽히 이해하고, 기기의 센서로부터 새로운 정보 소스를 포착하고, 딥러닝 역량을 구축하여 축적된 센서 데이터를 분석하는 기업일 것이다. 이 모든 것이 기업이 활동하는 데 통찰력을 주고, 기업의 물리적 인프라를 최대한 활용하는 데 도움을 준다. 우리는 전례 없는 수준으로 정확하게 의사 결정을 내려야 하는 시대에 있다. IoT 기술을 통해 세계 경제의 원동력이 될 새로운 차원의 생산성을 이끌어 낼 것이다.

## 2 IoT의 발전 역사

### 2-1 사물인터넷(IoT)이라는 용어의 등장

사물인터넷은 2018~19년도부터 시작해서 최근에서야 우리의 일상생활에 정착했다. 사물인터넷의 개념 자체는 최초의 전자기 전신이 설계된 1832년으로 시작되어 시대를

거슬러 올라간다. 전신(Telegraphy)은 다이렉트 통신을 가능하게 한 전기 신호 통신을 통한 기술이다. 하지만 본격적인 IoT 역사는 1960년대 후반 우리가 너무나도 당연하게 사용하고 있는 지금의 인터넷의 발명으로 시작되었으며 이후 몇십 년 동안 빠르게 발전했다. 1980년대에 들어 최초의 사물인터넷으로 연결된 장치는 미국 현지의 프로그래머가 제작한 코카콜라 자판기였다. 프로그래머들은 마이크로 스위치를 기계에 연결하고 그 당시 형태의 인터넷을 이용하여 음료를 충분히 차갑게 유지하는지, 일정한 수준의 냉각장치 작동이 들어가는지, 그리고 사용 가능한 코카콜라 캔이 있는지를 확인했다. 이 프로그래머의 발명을 토대로 전 세계는 상호 연결된 기기들의 개발부분과 현지 및 현장에서의 많은 연구와 실험을 촉진했다. 1990년 존 롬키(John Romkey)는 TCP, IP 프로토콜을 이용하여 토스터를 최초로 인터넷에 커넥트(Connect) 하였다. 그 이후로 1년이 지나 캠브리지 대학의 과학자들은 최초의 웹 카메라 프로토타입을 적용시켜 지역 컴퓨터실의 커피 포트에서 사용할 수 있는 커피의 양을 모니터링하는 아이디어를 발상하여 제작하기도 하였다. 그들은 분당 총 3회의 커피 포트의 사진을 찍을 수 있도록 웹캠을 프로그래밍한 후 이미지들을 로컬 컴퓨터로 전송시켜 모든 사람이 커피가 있는지 확인할 수 있도록 하였다. 1999년 비누, 샴푸, 칫솔 등 다양한 종류의 소비재를 제조 및 판매했던 P&G에서 브랜드 매니저로 근무하던 케빈 애시턴(Kevin Ashton)은 사물인터넷이라는 용어를 최초로 만들었으며, 이는 IoT 역사상 가장 중요한 연도이기도 하다. 애쉬턴은 유망하고 능력 있는 기술자로서 IoT 사물인터넷을 공급망 관리를 위한 RFID 태그의 기술을 기반으로 많은 장치를 커넥트하는 기술이라 언급하였다. 이는 자사의 제품들에 RFID 태그를 부착함으로써 제품들의 가시성을 확보할 수 있는 것처럼 세상에 존재하는 모든 사물이 서로 연결될 수 있다면 새로운 세상이 펼쳐질 것이라는 생각에서 기인했다.

## 2-2 사물인터넷의 진화

이후 사물인터넷의 개념은 RFID뿐만 아니라 다양한 센서 및 통신기술들과 결합하며 발전해 나가기 시작했다. 21세기에 접어들면서 2000년대에 사물인터넷이라는 단어는 갖은 매체와 미디어에서 사용되면서 가디언지, 포브스, 보스턴글로브와 같은 손에 꼽는 유명한 매체에서도 다루기 시작했다. 사물인터넷 기술에 대한 관심도가 꾸준히 증가하여 2008년도 스위스에선 사물인터넷 국제 콘퍼런스(Conference)가 개최되어 총 23개국에

이르는 많은 참가자들이 RFID, 단거리 무선통신 및 센서 네트워크에 의견을 나누었다. 또한 사물인터넷을 더욱 발전시키게 된 계기를 만드는 브랜드들의 제품 출시가 이어졌다. 그중 하나는 2000년도에 LG전자의 인터넷과 연결된 냉장고인데 사용자가 온라인 쇼핑과 더불어 화상 통화를 가능케 한 제품이다. 2010년대에 접어들어 사물인터넷은 2011년 네트워크 계층 프로토콜인 IPv6가 출시되면서 이후 일상생활 속 상호 연결이 되는 장치들이 널리 보급되어 전반적인 보편화를 이루게 되었다. 구글, 애플, 삼성, 및 제너럴 모터스와 같은 글로벌 대기업들은 자율 주행 자동차에서 스마트 글라스 등 IoT를 이용한 사물 장치의 생산에 중점을 두고 개발하였다. 이로 인해 우리의 삶 속에 제조뿐만 아니라 농업, 결제, 운송, 의료 등 다양한 분야에서 IoT는 모든 산업에 적용될 수 있다.

이처럼 빠른 발전 속도로 인해 IoT는 곧 전 세계에서 빠질 수 없는 기술이 될 것으로 예상된다. 예를 들어 자동차 시장에서 IoT는 최근 연평균 21% 성장률을 보이며, 해당 산업의 필수 기술임을 증명하고 있다. 이러한 IoT의 빠른 확장을 촉진시키는 요인들은 다음과 같다. 하락하는 센서 유지 비용, 데이터 수집과 세이브 비용 감소로 이어지는 클라우드 솔루션, 빠른 컴퓨팅 성능의 향상, 태블릿 PC와 스마트폰의 보급률 증가, 계속해서 확장되는 광범위한 인터넷 연결이 그 것이다. 이러한 IoT의 급속한 성장은 추후 우리가 살고 있는 세상을 근본적으로 바꿀 것으로 예상된다.

또한 IoT 기술은 별다른 큰 개발 없이도 블록체인, 인공지능, 빅데이터, 클라우드의 분야가 다른 기술과 결합할 때 더욱더 큰 장점으로 부각될 것이다. 하지만 IoT에도 단점이 존재하는데 보안상 취약점은 사이버 범죄자들의 공격에 개인정보가 유출될 수 있다. 이러한 우려사항은 IoT 기술이 추후 넘어야 할 장벽이다. 하지만 이러한 보안 문제에도 불구하고 상호 연결 기술이 빠르게 변화하고 진화하는 것은 우리가 더 나은 삶을 살고 일하는 방식을 개선하는 데 큰 도움을 줄 것이다.

### 2-3 USN(Ubiquitous Sensor Network)과 M2M(Machine-to-Machine)

USN이라고도 불리는 무선 센서 네트워크는 근거리 무선통신 기능을 포함하고 있는

소형의 센서 장치들이 결합하여 산불 감시, 하천 범람, 건물 내 온도 분포 등 특정 장소의 상태 및 환경 변화 정보를 종합적으로 수집하여 관리하기 위한 기술이다.

반면, M2M 기술은 이동통신기술을 이용하여 멀리 떨어져 있는 기계장치를 연결함으로써 효율적으로 장치를 운용하는 기술이다. 병원에서는 응급상황, 환자의 상태 모니터링, 의학 데이터 등을 연결하여 건강관리 시스템을 구축하기도 한다. 은행의 현금지급기(ATM), 택시에 설치된 카드 결제기 또한 이러한 M2M 기술의 대표적인 예에 해당한다.

## 2-4 사물인터넷의 본격화

산불감시나 현금지급기 등 주로 개별적인 목적을 위해 사용되던 무선 센서 네트워크와 M2M 기술은 활용 분야가 한정적이었으며 도입 비용도 높아서 그다지 성공적이지 못했다. 이후 다양한 센서장치와 디바이스를 결합하여 새로운 사용자 가치를 창출하려는 시도가 나타나면서, 소비자 유통, 헬스케어, 스마트홈 등 다양한 분야에서 사물인터넷 시대가 본격화되기 시작했다.

## 3 4차 산업혁명과 IoT

현대에는 4차 산업혁명이 주목받고 있다. 특히 4차 산업혁명에서 주목받는 기술은 인공지능으로써, 3차 산업과 4차 산업을 구분하는 핵심기술이다. 그러므로 4차 산업혁명과 인공지능은 분리해서 생각할 수 없다. 그러나 인공지능 기술이 4차 산업혁명의 전부는 아니다. 인공지능 서비스에 변화를 가져오는 것은 맞지만, 인공지능 기술에만 초점을 뒀서는 안 된다. 인공지능과 함께 사물인터넷도 중요하다. 인공지능은 사람의 두뇌 역할을 한다면, 사물인터넷은 사람의 눈, 코, 입, 귀와 같은 역할을 한다. 그러므로 인공지능을 실현시키기 위해서 사물인터넷도 함께 고려해야 한다. 그래서 이 두 기술에 대한 내용과 연관성을 파악할 필요가 있다.

### 3-1 4차 산업혁명의 특징

4차 산업혁명에 대한 특징을 살펴보기 이전에, 지금까지 발생한 산업혁명에 대해서 우선 살펴보자.

〈표 III-1〉 단계별로 산업혁명을 정리한 비교표

구분	1차 산업혁명	2차 산업혁명	3차 산업혁명	4차 산업혁명
시기	18세기 후반	19~20세기 초	20세기 후반	2016년~
핵심기술	열에너지와 증기기관	전기 및 벨트컨베이어	인터넷 기반의 지식정보	사물인터넷, 인공지능
핵심가치	생산의 자동화	대량생산	정보	편리성
주요 특징	열과 물의 힘을 이용한 인력의 기계대체화	전기에너지를 이용한 대량생산 체계 구축	정보 대량 확산으로 인한 합리성 증가	기기 지능화로 인한 편리성 증대

출처: 한국정보기술학회지(2017)

최초의 산업혁명은 18세기 후반에 나타났다. 이를 1차 산업혁명이라고 하는데, 인력이 기계로 대체되는 시기이다. 열 에너지가 등장해서 부분적으로 자동화가 가능했는데, 생산성은 섬유분야 외에는 높지 않았다. 1차 산업혁명 시대에는 증기기관 기차가 처음으로 등장했다. 따라서 기차에 필요한 인프라가 구축되었는데, 다리·터널·항만 등의 기반시설이 이 시기에 설치되었다고 볼 수 있다.

2차 산업혁명은 ‘전기’가 처음으로 등장한 시기이다. 2차 산업혁명은 19세기 후반에 발생한 혁명인데, 공장에 전기를 공급하면서 컨베이어 벨트도 도입됐다. 이로 인해서 대량생산이 촉발되었다. 참고로 이 시기를 영국에서 발생한 ‘산업혁명’ 시기로 볼 수 있다. 이 시대의 공장 경영원칙은 ‘효율적인 생산관리’이다. 대량생산과 비용 최소화가 공장에서는 큰 관심사였다. 그래서 이때부터 프레드릭 테일러(Frederick W Taylor)의 ‘과학적 관리론’이라는 이론이 등장했는데, 이 개념이 경영의 시초가 된다.

3차 산업혁명은 우리가 이미 잘 알고 있듯이 ‘정보혁명’으로 정의할 수 있다. 인터넷기술이 등장해서 전 세계의 모든 정보의 공유가 가능해진 시기이다. 1969년 미국 국방성이 4개의 대학교에서 사용하는 사설 네트워크를 연결하는 과제를 시초로, 전 세계를 네트워크로

연결하는 단계로 발전하면서 지금의 인터넷이 등장했다고 볼 수 있다. 2차 산업혁명의 중심은 대량생산의 주체인 ‘공급자’에 있었다. 그러나 3차 산업혁명은 중심을 ‘공급자’에서 ‘수요자’로 전환시켰는데, 이는 상품에 대한 정보 공유가 가능해져 기업의 수익에 영향을 미쳤기 때문이다. 3차 산업혁명의 특징은 정보 공유로 인한 ‘효율성의 극대화’이다. 그리고 기업에서는 양적인 측면보다는 질적인 측면에 관심을 가지게 했다

그렇다면 4차 산업혁명은 어떻게 발생했고 특징은 무엇일까? 4차 산업혁명은 대량의 정보 때문에 발생한 것으로 볼 수 있다. 그러나 이러한 정보들은 대부분 질적인 부분이 많아서, 분석하기에는 적절하지 못했다. 그런데 방대한 질적인 정보를 분석할 수 있는 빅데이터 기술이 등장하며 상황이 바뀌게 된다. 이때부터가 본격적인 4차 산업혁명의 시작으로 볼 수 있다. 이러한 분석기술은 점점 더 고도화되며 인공지능의 탄생에 영향을 미쳤다. 인공지능은 데이터를 더욱 정교하게 분석할 뿐만 아니라 산업의 자동화를 이루고 있다. 이와 더불어 정보 소비와 생산주체가 3차 산업혁명 시대에 사람에 국한됐다면, 4차 산업혁명시대에는 모든 사물이 정보 소비와 생산주체가 가능하다. 이는 사물인터넷으로 인해서 도래했다.

이처럼 인공지능 기반으로 모든 것들이 연결된 사회를 ‘4차 산업혁명’으로 볼 수 있다. 4차 산업혁명은 3차 산업혁명과 비교 시, 정보의 질적인 측면과 공유 범위가 넓어진 셈이다. 여기서 참고로 인공지능 기반으로 정보를 가공해서 사용자에게 제공하는 정보를 능동형 정보(Actionable Intelligence)라고 한다. 능동형 정보는 경영전략에서 시초가 돼, 빅데이터 등장과 함께 사용되고 있는 용어이다. 능동형 정보는 정보를 받는 대상에게 통찰력을 제공해서 더 나은 결과를 이끌 수 있게 하는 정보를 말한다. 다시 말해 단순 사실적인 정보가 아닌, 스스로 분석해서 알려주는 정보를 말한다. 에너지 분석 시스템을 예로 들어보겠다. 3차 산업혁명 시대에는 환경·에너지 사용량 정보를 제공했다면, 4차 산업혁명은 두 요인의 관계를 분석해서 사용자가 적절한 해법을 제시할 수 있게 한다. 이때 두 요인에 대한 분석 기반이 되는 것은 빅데이터 혹은 인공지능이다.

### 3-2 4차 산업혁명과 IoT-AI 플랫폼

위의 산업혁명의 단계마다 확인할 수 있는 사실은, 산업혁명을 발생시키기 위한 ‘촉발 기술’이 있었고 이에 따른 결과가 있었다는 것이다. 1차 산업혁명은 ‘열에너지’, 2차 산업혁명은 ‘전기’, 3차 산업혁명은 ‘인터넷’이 촉발 기술의 역할을 했다. 아울러 이러한 기술에 의해서 1차 산업혁명은 생산의 자동화, 2차 산업혁명은 대량생산 그리고 3차 산업혁명은 ‘합리성 및 효율성 향상’이라는 결과를 낳았다.

그렇다면 4차 산업혁명의 촉발기술은 무엇이고, 4차 산업혁명 시대에는 어떠한 변화를 기대할 수 있을까? 앞서 언급했듯이 4차 산업혁명은 수많은 질적 정보로 인해서 발생한 산업혁명이라고 말했다. 그러나 이러한 정보들을 분석할 수 없으면 아무런 의미가 없다. 그래서 3차 산업혁명 시대에는 이러한 정보들은 가치 없이 공유되었다. 그러나 빅데이터 기술이 등장해서 이러한 정보들의 분석이 가능케 되었다. 이로 인해서 기존에 제공하지 못했던 서비스 구현이 가능하게 됐고, 더욱더 고도화돼 ‘인공지능’ 기술도 급격히 발전하기 시작했다. 결국 4차 산업혁명의 촉진기술은 인공지능으로 볼 수 있다. 그런데 인공지능의 핵심은 자동화이다. 다시 말해 사람의 일을 대신하는 것이 인공지능의 핵심이다. 이를 위해서는 기기들 간에도 정보를 공유해서 스스로 판단할 수 있게 하는 매개체가 필요하다. 그래서 뒷받침된 기술이 사물인터넷이다. 수많은 사물 센서들에서 발생하는 데이터를 사물인터넷 네트워크 인프라를 통해 전송한다. 네트워크 인프라는 4차 산업혁명, Big Data 시대를 이끌어가는 핵심 인프라의 하나로 수많은 사물들을 가벼운 저용량의 배터리로 연결한다.

〈그림 III -2〉 IoT 네트워크 요구 사항



사물인터넷은 인공지능 시스템에 정보를 제공해줌과 동시에, 인공지능이 제공하는 분석 정보들을 받아서 사용자에게 제공하거나, 자동으로 처리하기도 한다. 결론적으로 4차 산업혁명의 촉발 기술은 사물인터넷과 인공지능으로 정의할 수 있다. 그렇다면 이 둘의 관계로 인해서, 어떤 변화를 일으킨 것일까? 바로 사물을 활용한 '스마트 서비스' 구현이다. 이 두 기술의 결합은 스마트 서비스를 구현 가능케 했는데, 이는 사용자 편리로 이어지게 된 것이다.

사이버 보안 산업을 예로 들겠다. 기존 보안 솔루션은 악성공격으로 의심되는 사건들을 팝업형태로 모두 관리자에게 보고했다. 관리자는 이 모든 내용들을 다 보아야 했기 때문에, 업무 부담이 많았다. 그러나 현재 스마트 기술 기반의 보안 솔루션은 다르다. 보안 솔루션은 모든 사건들을 관리자에게 알려주는 것이 아니라 확실한 것들을 알려준다. 그뿐만 아니라 사건의 연관관계를 알려주어 관리자가 좀 더 쉽게 분석할 수 있도록 보안 서비스를 제공해준다. 결국 이는 관리자의 편리성을 증대시키는 역할을 한다. 그뿐만 아니라 업무 성과에도 기여한다. 이처럼 보안 솔루션의 지능적인 역할은 관리자에게 매우 중요하기 때문에 인공지능 주요 업체인 IBM이 개발한 시스템도 보안 솔루션에 도입되어 활용되는 중이다.



## 1 5가지 계층으로 이뤄진 IoT 프레임워크

센서 → 연결 → 수집(클라우드) → 분석(빅데이터) → 실행(앱 등)

사람들은 사물인터넷의 일반적인 정의가 '장님과 코끼리' 동화에 나오는 코끼리 같다고 한다. 사람들이 눈을 가리고 각자 코끼리의 다른 부분을 만져본 뒤 자신이 만져본 경험을 바탕으로 코끼리를 설명하는 것이 마치 여러 분야의 IT 전문가가 현재의 사물인터넷을 정의하는 것과 유사하기 때문이다. 이러한 시각 차이로 인해 사물인터넷의 정의와 개념은 여러 가지일 수 있다. 여기에서는 사물인터넷의 주요 구성 요소를 모두 아우르는 프레임워크를 제시하고자 한다. 기술적인 측면과 비즈니스 측면에서 모두 사용할 수 있는, 다섯 가지 구조층으로 이루어진 모든 사물인터넷 솔루션에 적용 가능한 프레임워크이다.

IoT 애플리케이션을 구현하기 위한 기술을 구축 · 구매 · 판매 · 투자할 때는 해당 산업의 다양한 구성 요소 또는 부분을 이해하는 데 사용할 수 있는 프레임워크를 개발하는 것이 중요하다. 또한, 이 프레임워크를 비즈니스 및 기술 담당자가 이해해야 한다. 마지막으로 사물인터넷 솔루션을 제공하고자 하는 기업은 이를 개발하기 위한 프레임워크를 소유하고, 사용하는 것이 중요하다.

먼저 IoT 프레임워크의 다양한 부분을 정의하고자 한다. IoT 프레임워크는 센서, 연결, 수집(클라우드), 분석(빅데이터), 실행(앱)이라는 다섯 가지 주요 계층으로 구성된다.

첫 번째 계층은 센서다. 스마트폰, 스마트워치 등의 사물에는 여러 종류의 센서가 장착되어 있고, 이러한 센서들은 여러 가지 네트워크 통신 방법으로 인터넷에 연결된다. 일단 ‘연결’된 다음, 데이터를 클라우드나 온프레미스(On-premise) 서버로 수집하도록 설계된 기술이 ‘수집’이다. 수집된 데이터는 대부분 시간, 분 또는 초 단위로 전송되는 시계열 데이터(Time-Series Data)이다. 네 번째 계층은 ‘분석(빅데이터)’이다. 사용자가 뭔가를 입력하도록 유도해야 하는 IoP 애플리케이션의 세계와 달리, IoT 애플리케이션은 끊임없이 데이터를 얻는다. 마지막으로 ‘실행’ 계층은 소프트웨어 애플리케이션 기술과 비즈니스 모델을 모두 포함한다. 여기에서 비즈니스 모델은 ‘센서’를 장착한 사물을 만드는 회사와 사물을 이용하여 의료, 운송, 건설 등의 서비스를 제공하는 회사의 비즈니스 모델을 말한다.

## 1-1 센서

스마트워치, 냉장고, 심지어 농장의 트랙터 등의 사물은 각종 ‘센서’를 장착하여 더욱 스마트해지고 연결성이 높아지고 있다. 차세대 기계를 구축하거나 구입하려면 센서, CPU 아키텍처, 운영 체제, 패키징, 보안을 고려해야 한다. 센서는 무어의 법칙에 따라 매년 생산 비용이 크게 떨어지고 있다. 더 많은 센서가 저가 컴퓨터에 장착되고 있으며, 저가 컴퓨터의 종류는 단순한 마이크로컨트롤러(Microcontroller)부터 ARM 또는 Intel 명령어 세트 아키텍처(Instruction Set Architecture)를 지원하는 완전한 기능의 CPU에 이르기까지 다양하다. 더 강력한 소프트웨어를 지원하려면 보다 강력한 프로세서를 사용해야 한다. 또한, 이와 함께 소프트웨어의 보안은 해킹 등의 위험으로 점점 더 그 중요성이 대두되고 있다.

## 1-2 연결

사물은 다양한 방법으로 인터넷에 연결될 수 있다. 사물을 연결하는 데는 얼마나 많은 데이터를 전송해야 하는지, 얼마나 멀리 데이터를 전송해야 하는지, 얼마나 많은 전력 배터리량을 보유하고 있는지에 따라 다양한 기술이 필요하다. 또한, 연결을 관리하는 방법과 보호·보안 방법에 대해 상위 수준의 더 많은 옵션이 있다.

### 1-3 수집

각종 센서를 장착한 사물은 엄청난 양의 데이터를 생성할 수 있으며, 이렇게 방대한 양의 데이터, 즉 빅데이터는 기하급수적으로 커질 것으로 예상된다. 이는 센서 기술의 발달과 가격 하락, 그리고 사물인터넷 통신기술 등의 발전 때문이다. 사물이 발생하는 이러한 데이터를 SQL, NoSQL, 전통적인 시계열 및 차세대 시계열 수집아키텍처를 사용하여 클라우드 또는 온프레미스 서버에 수집하고 저장할 수 있다.

### 1-4 분석

사물에서 오는 데이터가 점점 더 많아지므로 그 데이터에 기술을 적용해 분석해야 한다. 분석 계층에서 사용하는 제품에는 쿼리 기술과 지도(Supervised) 및 비지도(Unsupervised) 머신러닝 기술이 모두 포함된다. 지금까지의 산업에서는 주로 사람에 대한 데이터를 분석하는데 중점을 두었다. IoT에서는 기존 데이터 분석 기술이 사물에 대한 데이터를 분석하는 것으로 혁신적인 발전을 이룰 것으로 기대된다.

### 1-5 실행

IoT 애플리케이션과 마찬가지로, IoT 애플리케이션을 구축하기 위한 패키지 애플리케이션(예: ERP, CRM)과 미들웨어가 있다. 애플리케이션을 구입했던 구축했던 결국은 비즈니스 성과를 이끌어 내야만 한다. 각종 센서를 장착한 사물이 점차 복잡해지고 소프트웨어에 의해 활성화되면서, 소프트웨어 유지 보수 및 서비스에서 만들어낸 혁신이 사물 서비스에도 적용될 것이다. 소프트웨어를 서비스 형태로 제공하는 SaaS(Software-as-a-Service)가 업계에 혁명을 가져왔던 것처럼 사물도 서비스의 형태로 제공함으로써 혁신을 이룰 가능성이 있다.

## 2 IoT 계층별 기술, 기능 및 특징

위에서 언급했듯이 사물인터넷은 '센서 → 연결 → 수집 → 분석 → 실행'의 5개 계층으로 이뤄진다. 여기에서는 센서와 연결 기술에 대해 자세히 알아보자.

### 2-1 센서

센서는 무어의 법칙을 따르기 시작해 매년 큰 폭으로 가격이 하락하고 있다. 센서는 저렴한 컴퓨터에 더 많이 부착되며, 저가 컴퓨터의 종류는 단순한 마이크로컨트롤러부터 ARM 또는 Intel 명령어 세트(Instruction Set) 아키텍처를 지원하는 CPU에 이르기까지 다양하다. 더 강력한 프로세서를 사용할수록 더 강력한 소프트웨어를 지원할 수 있다.

#### 가. 센서 종류

센서는 수백 가지가 있는데, 더 구체적으로 설명하기 위해 휴대전화를 살펴보자. 매우 일반적으로 사용되는 센서 중 하나인 가속도계는 휴대전화에 내장되어 휴대전화가 자유낙하 대비 겪고 있는 가속도를 측정한다. 이 센서는 세 축을 따라 휴대전화 기기의 방향을 측정하는 데도 사용된다. 휴대전화의 애플리케이션(이하 앱)은 이 데이터를 사용하여 휴대전화의 방향이 세로인지 가로인지, 화면이 위를 향하는지 아래를 향하는지 확인한다. 자이로스코프는 또 다른 센서로 가속도계보다 더 정밀하게 방향 정보를 제공한다. 휴대전화가 얼마만큼 회전하고, 어느 방향으로 움직이는지 알려준다.

또 다른 센서는 자력계이다. 자력계는 자기장의 강도와 방향을 측정한다. 나침반 앱과 금속 탐지용 앱에서 사용된다. 근접 센서는 물체에 적외선 광선을 비추고, 반사된 적외선 광선을 탐지기가 수집하는 방식으로 작동한다. 수화기 근처에 내장되어 사용자가 통화 중인지를 감지하고 화면을 끄도록 알려주어 배터리 소모를 줄여준다.

또한, 주변 빛의 밝기를 측정하는 광센서가 있다. 휴대전화는 광센서 데이터를 사용하여 디스플레이의 밝기를 자동으로 조정한다. 일부 휴대전화에는 대기압을

측정하는 기압계가 내장되어 있다. 기압계는 해발 고도를 결정하는 데 사용되므로 GPS 정확도가 향상된다. 모든 휴대전화에는 온도계가 있고, 온도계가 하나 이상인 경우도 있다. 전화기의 구성 부품이 과열되었다고 감지되면 손상을 막기 위해 시스템을 자체 종료한다. 삼성의 특정 갤럭시 모델은 휴대전화가 공기 중 습도 센서를 사용하기도 한다. 습도계가 제공한 데이터는 앱에서 사용되어 사용자가 '최적의 기온과 습도 영역'에 있는지 알려 준다.

보수계는 사용자의 걸음 수를 세주는 센서이다. 걸음 수를 측정할 때 일부 휴대전화는 가속도계의 데이터를 사용하지만, 전용 보수계가 더 정확하고 전력 면에서 효율적이다. 구글 넥서스 5는 전용 보수계가 내장된 몇 안 되는 휴대전화 중 하나이다. 갤럭시 X5 같은 일부 휴대전화에는 심장 박동 모니터가 있으며, 손가락 혈관의 미세한 맥동을 감지하여 구현된다. 또한, 대부분의 애플 사용자는 잠금 화면 암호 대신 지문 인식 센서를 사용한다. 마지막으로 보통 사람들이 휴대전화에 내장되어 있다고는 예상치 못할 센서로 방사선 센서가 있다. 사프 팬톤 5에는 해당 지역의 현재 방사능 레벨을 측정하는 앱이 있다. 마이크와 카메라까지 센서 목록에 추가하면 휴대전화 한 대에 적어도 14개의 서로 다른 센서가 내장되어 있다.

#### 나. 컴퓨터 아키텍처

모든 스마트 커넥티드 사물 또는 기계에는 일부 소프트웨어를 실행할 수 있는 일종의 CPU(Central Processing Unit)가 내장되어 있다. 이러한 사물은 전력 소비, 물리적 크기, 비용 등에 제한을 받기 때문에, CPU의 기본 아키텍처를 구성함에 있어서 여러 가지 균형적 요소(tradeoff)가 존재한다.

마이크로컨트롤러는 가장 기본적인 CPU의 형태로 기억장치(Memory Storage)에 대한 접근이 제한적이고, 일반적으로 전력 소모가 적은 간단한 명령어 세트(Instruction Set)를 가지고 있다. 마이크로컨트롤러는 여러 가지 등급이 있는데, 가장 기본 등급의 예로 인텔이 제조한 8비트, ATmega 시리즈 마이크로컨트롤러를 사용하는 아두이노 등이 있다. 사물은 이렇게 더 작고 저렴하며 기능이 적은 마이크로컨트롤러를 사용할 수 있다. 다음 단계는 ARM 또는 인텔이 제공하는 모든 기능의 명령어 세트를 갖춘 마이크로컨트롤러이다.

ARM은 축소 명령 집합 컴퓨터(RISC; Reduced Instruction Set Computer)로 대부분의 개인용 컴퓨터에서 일반적으로 쓰이는 Intel x86 프로세서보다 명령어 세트를 구현하는 데 필요한 트랜지스터가 훨씬 적다. 결과적으로 이 접근법은 비용, 발열 및 전력 사용량을 감소시킬 수 있다. 라즈베리 파이는 32비트 ARM 프로세서를 기반으로 한다. 최근 휴대폰과 태블릿을 지원하는 애플 A7은 애플이 설계한 64비트 시스템의 ARM CPU이다. 많은 기업이 ARM을 사용하지만 인텔의 x86 아키텍처는 여전히 많은 서버와 랩톱을 구동하는 데 쓰이며, 이와 관련된 소프트웨어 개발 도구가 많다. 인텔 아톰(Atom)은 인텔의 초저전력 마이크로프로세서 제품군의 브랜드 이름이다.

#### 다. 소프트웨어

특정 사물에 소프트웨어를 실행시키는 운영 체제는 여러 가지가 있다. 무엇을 선택하느냐는 소프트웨어에 필요한 메모리 공간, 개발 환경 및 실시간 요구 사항에 따라 달라진다.

##### 1) 메모리 사용 공간

컴퓨터에서 실행하는 프로그램의 메모리 사용 공간은 프로그램이 실행되는 동안의 런타임 메모리 요구 사항을 나타낸다. 큰 프로그램일수록 더 큰 메모리 사용 공간이 필요하다. 하지만 메모리 사용 공간의 가장 큰 부분이 소프트웨어 프로그램 자체가 아닌 경우도 많다. 오히려 런타임(run-time) 환경에 의해 도입된 구조가 메모리 사용 공간을 증가시킨다. 예를 들어 자바(Java) 프로그램에서 메모리 사용 공간은 주로 JVM(Java Virtual Machine) 런타임 환경으로 구성된다.

##### 2) 소프트웨어 개발 환경

IDE(Integrated Development Environment)라고도 부르는 소프트웨어 개발 환경은 소프트웨어 개발을 위해 컴퓨터 프로그래머에게 포괄적인 기능을 제공하는 소프트웨어이다. 소프트웨어 개발 환경은 일반적으로 소스 코드 편집기, 빌드 자동화 도구, 디버거로 구성된다. 마이크로컨트롤러 프로그래밍을 위해 아두이노 플랫폼은 C, C++, 자바 프로그래밍 언어를 지원하는 IDE를 제공한다. 기업용 사물을 생산하는 기업은 프로그래머에게 어떤 IDE가 필요한지 고려해야 한다.

### 3) 운영 체제

운영 체제 또는 런타임 환경에는 많은 애플리케이션에서 일반적으로 사용하는 소프트웨어가 들어 있다. IoT 애플리케이션 세계에는 리눅스, 윈도 등 익숙한 운영 체제가 있다. IoT 애플리케이션에는 이른바 '실시간' 운영 체제가 필요하다.

실시간 운영 체제(RTOS; Real-time Operating System)는 시간 임계(time-critical) 애플리케이션 실행을 목적으로 하는 운영 체제이다. 시간적 요소가 중요하기 때문에 처리 시간 요구 사항(OS 지연 포함)은 10분의 1초나 더 짧은 시간 단위로 측정된다. RTOS의 핵심 특징은 애플리케이션 작업을 수락하고 완료하는 데 걸리는 시간의 일관성이다. 실시간 OS의 주요 요소는 최소한의 인터럽트 지연시간과 스레드 전환 대기시간이다. 실시간 OS는 주어진 시간 동안 수행할 수 있는 작업량보다는 얼마나 빨리 또는 얼마나 예측 가능하게 응답할 수 있는지를 더 많이 평가한다. 윈드 리버社は 실시간 처리를 지원하는 데 일반적으로 사용하는 RTOS를 제공하는 회사이다. 물론 프로세서가 더 빨라지고 스마트 기계의 수요가 다양해질수록 RTOS가 아닌 기존 운영 체제를 더 많이 사용할 수도 있다.

### 4) 패키징

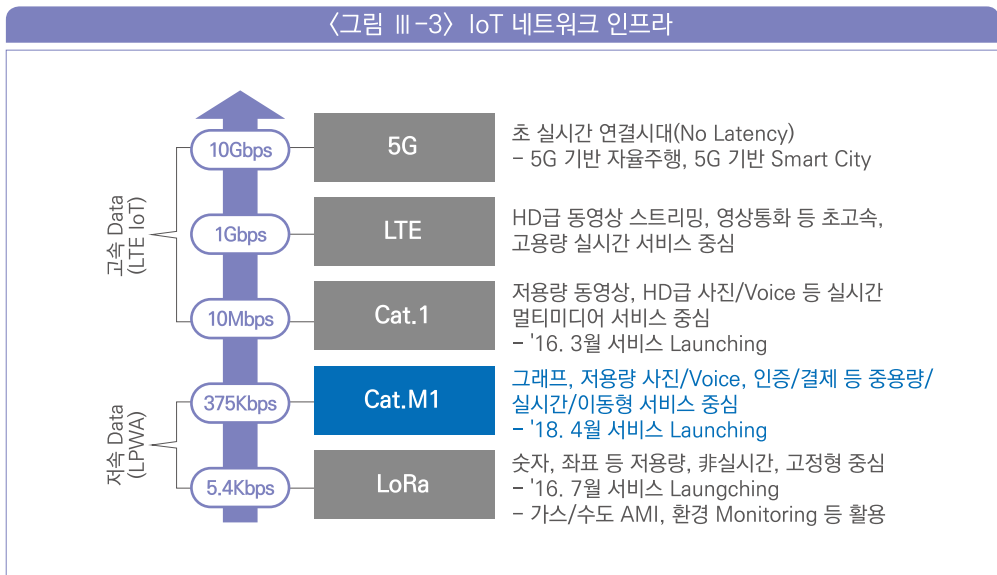
패키징(packaging)은 전자 공학의 주요 분야이며 다양한 관련 기술을 포함한다. 패키징에 있어 기계적 손상, 냉각, 무선 주파수 소음 방출, 정전기 방전으로부터의 보호는 필수 고려 사항이다. 소량으로 제조된 산업 장비는 카드 케이지(card cage) 또는 조립식 박스 같은 표준화된 시판용 케이스(Enclosure)를 사용할 수 있다. 대중용 소비자 기기는 소비자에게 호소력을 높이기 위해 고도로 전문화된 패키징을 이용할 수 있다.

대량으로 사용되는 패키징 기술 중 하나는 SoC(System-on-a-Chip)이다. SoC는 컴퓨터 또는 다른 전자 시스템의 모든 구성 요소를 단일 반도체 칩으로 결합한 집적 회로이다. 디지털, 아날로그, 혼합 신호, 무선 주파수 기능을 포함할 수 있다. SoC와 마이크로컨트롤러의 차이점은 메모리량이다. 마이크로컨트롤러는 일반적으로 메모리가

100KB 미만인 반면, SoC는 더 큰 메모리 사용 공간을 필요로 하는 소프트웨어(예: 리눅스)를 실행할 수 있는, 더 강력한 프로세서에 사용된다. 특정 애플리케이션을 위한 SoC를 구성하는 것이 타당하지 않을 경우 패키지 하나에 여러 칩을 포함하는 SiP(System-in-Package)가 대안이 될 수 있다. SiP보다 SoC가 제조 수율이 높고 패키징이 간단하기 때문에 대량생산 시 SoC가 비용 효율성이 높다고 여겨진다.

## 2-2 연결

IoT 네트워크 인프라는 지속적으로 새로운 기술이 개발되고 있다. 활용 용도와 데이터 속도 등에 따라 LoRa, LTE, 5G 등 다양한 통신기술이 존재한다.



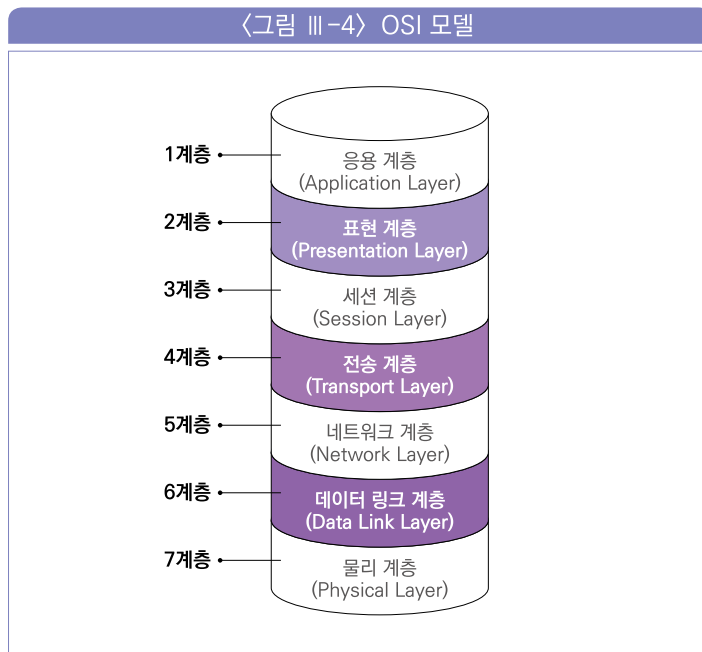
사물을 연결하려면 전송해야 하는 데이터의 양, 전송해야 할 거리, 보유하고 있는 전력량에 따라 다양한 기술을 조합해야 한다. 또한, 연결을 관리 · 보호 · 보안하는 방법으로 상위



수준에서 많은 선택권이 있다. 여기에서는 기계 연결과 몇 가지 기본 원칙에 대해 간단히 설명하고자 한다.

### 가. 네트워킹의 기본

네트워킹 분야에서 OSI 스택(네트워크 통신에서 생기는 여러 가지 충돌 문제를 완화하기 위해 국제표준기구(ISO)에서 표준화된 네트워크 구조를 제시한 기본 모델. 통신망을 통한 상호 접속에 필요한 제반 통신 절차를 정의하고 이 가운데 비슷한 기능을 제공하는 모듈을 동일 계층으로 분할하여 총 일곱 계층으로 이뤄진다)에 대한 내용은 빠질 수 없다. OSI(Open Systems Interconnection) 스택은 기본 내부 구조나 기술에 관계없이 원격 통신 또는 컴퓨팅 시스템의 통신 기능을 특성화하고 표준화하는 개념적 모델이다. OSI의 목표는 표준 프로토콜을 활용하여 다양한 통신 시스템의 상호 운용성을 이루는 것이다. OSI 모델은 통신 시스템을 일곱 개의 추상화 계층으로 분할하여 정의한다.



출처: Imperva.com











각 계층은 상위 계층에 서비스를 제공하는 형식으로 동작하며, 또한 하위 계층으로부터 서비스를 제공받는다. 예를 들어 네트워크를 통해 오류 없는 통신을 제공하는 계층은 상위 계층의 애플리케이션에 필요한 경로를 제공하는 한편, 하위 계층을 호출하여 해당 경로의 내용을 구성하는 패킷을 보내고 받는다. 이러한 스택 구조는 물리층의 단순한 전자적 레벨에서 응용 레벨까지 다양한 연결 수준을 이해하는데 유용한 프레임워크이다.

데이터 링크층(OSI; Data Link Layer)은 노드 간 데이터 전송을 제공하며, 여기에서 데이터 링크란 물리적으로 직접 연결된 두 노드(Node) 간의 링크를 말한다. 이 계층은 물리층에서 발생할 수 있는 오류를 감지하고 수정한다. 물리적으로 연결된 두 장치 간의 연결을 설정하고 종료하는 프로토콜을 정의하고, 기기들 사이의 데이터 흐름을 제어하는 프로토콜을 정의한다. 데이터 링크층은 네트워크 장치의 데이터에 대한 액세스 권한과 전송 권한을 제어하는 역할을 한다. 또한, 네트워크층 프로토콜을 식별하고 캡슐화하며 오류 검사와 패킷 동기화를 제어한다. 이더넷, 와이파이, 지그비 등의 연결 기술은 모두 데이터 링크층에서 작동한다.

#### 나. 범위 대 전력 사용량

사물은 어떤 장소에 놓거나 이동할 수 있다는 특성 때문에 물리적으로 인터넷에 연결시킬 수도 있지만, 대부분 특정 형태의 무선 기술이 필요하다. 와이파이, 3G, 4G, 5G, 지그비, NFC, 로라WAN, 위성 등 수많은 전략적 선택 항목이 있다. 어떤 전략을 선택하느냐는 전력 소비, 범위 요구 사항, 데이터 전송 속도, 비용, 안테나 크기, 환경과 관련된 몇 가지 기본 원칙에 따라 달라진다.

〈표 III-2〉 범위 대 전력 사용량

	WiFi	ZigBee	Bluetooth	NFC
POWER	 High	 Low	 Classic: Mid   LE/Smart: Low	 Tag: Zero   Reader: Very low
TANGE	 30-100m	 10-20m	 10m	 < 0.1m

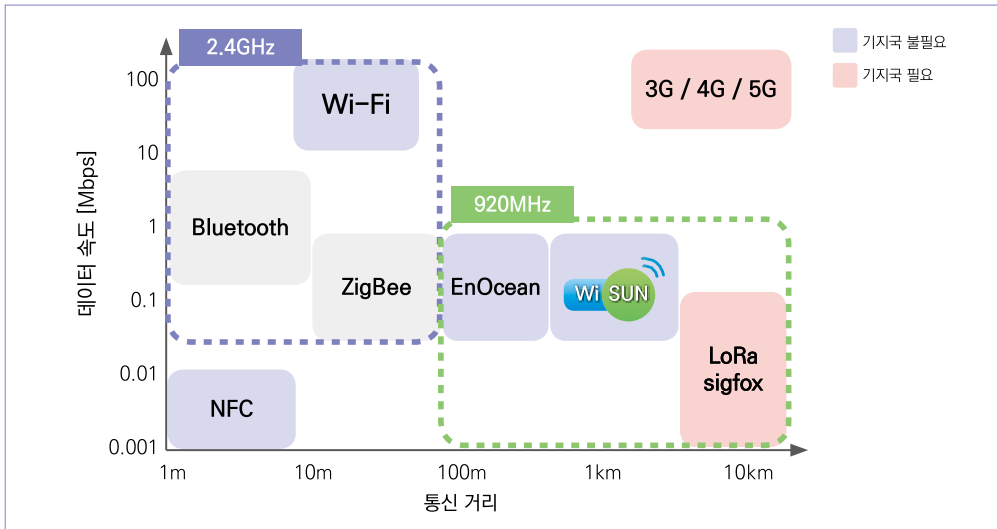
출처: 티모시 추, 『IoT 비즈니스 전략』, 길벗(2020)

위의 표는 네 가지 연결 기술의 범위와 전력 소비 사이의 균형을 보여준다. 30~100미터 범위의 와이파이의 최대 범위가 10미터인 블루투스나 0.1미터 미만인 NFC보다 훨씬 많은 전력이 필요하다. 많은 사람이 휴대전화의 배터리 전원을 절약하려면 와이파이를 꺼야 한다는 점을 알고 있다. 일반적으로 무선신호는 거리의 제곱으로 감소한다. 다시 말해 범위를 두 배로 늘리려면 전력 소비는 네 배 증가해야 한다. 즉, 배터리가 더 커야 한다.

#### 다. 범위 대 데이터 전송 속도

또 다른 절충 사항은 범위 대 속도이다. 주파수가 높아지면 사용 가능한 대역폭은 일반적으로 증가하지만, 장애물을 극복하기 위한 거리와 능력은 감소한다. 거리와 상관없이 2.4GHz 장치는 900MHz에 비해 추가 경로 손실이 8.5dB 발생(3dB는 50% 손실)한다.

〈그림 III-5〉 범위 대 데이터 전송 속도



출처: Rohm Semiconductor

따라서 주파수가 높을수록 대역폭 성능이 높아지지만, 같은 범위를 달성하려면 더 많은 전력이 필요하다. 낮은 주파수는 대역폭이 낮지만, 더 먼 거리에 이를 수 있다. 그러나 낮은 주파수가 높은 주파수와 동일한 범위를 달성하려면 더 큰 안테나가 필요하다.

〈그림 III-6〉 안테나 크기



출처: 티모시 추, 『IoT 비즈니스 전략』, 길벗(2020)

비 오는 날 위성 TV를 시청해본 사람이라면 누구나 알고 있겠지만, 주변 환경도 네트워크 성능에 영향을 준다. 제조업체는 가시선(Line of Sight) 범위 수치를 광고한다. 가시선이란 안테나 A와 안테나 B 사이에 장애물이 없는 상태로 하나의 안테나에서 반대편 안테나를 눈으로 볼 수 있음을 의미한다. 만약 안테나 B가 건물 내부에 위치해 안테나 A에서 안테나 B를 볼 수 없다면 가시선으로 간주하지 않는다. 경로의 모든 장애물은 각 장애물 유형에 따라 가시선 범위 수치를 감소시킨다. 장애물의 유형, 위치, 수는 모두 경로 손실에 영향을 준다.

## 1 IoT 활용에 따른 설계

### 1-1 IoT 센서 활용과 설계

#### 가. 센서 시장 동향

스마트 시티, 자율 주행과 5G 등 사물과 사물이 직접 상호작용하는 IoT 관련 부품수요 증가세에 맞추어 IoT 센서의 수요도 크게 늘어날 것으로 전망되고 있다. IoT 센서 시장은 미국(33.7%), EU(19.4%), 아시아·태평양(32.3%)로 미국과 아시아·태평양 시장이 양분하고 이어서 EU가 시장을 점유하고 있는 상황이다. 미국 지역은 미국이, 아시아 지역은 일본이, EU 지역은 독일이 IoT 센서 시장을 과점하며 전 세계 산업을 주도하고 있다.

#### 나. IoT 센서 시장 변화

글로벌 IoT 센서는 신규 시장(전통 제조업, 농업 등)의 부상, 유사 제조공정에서의 기능 통합, 부품과 솔루션의 결합이라는 3대 트렌드의 변화가 발생하고 있다.

〈표 III-3〉 지역별 시장 점유율 및 전망

(단위: 10억 US\$)

지역	'20	'21	'23	'25	CAGR('20~'25)
미국	12.5	13.9	20.6	31.1	20.0%
EU	8.2	10.0	7.9	18.5	17.7%
아시아 태평양	12.0	13.5	20.1	31.2	20.9%
기타	3.9	3.5	5.3	6.8	12.1%
합계	36.6	40.9	53.9	87.6	19.1%

출처: Markets and Markets(2019)

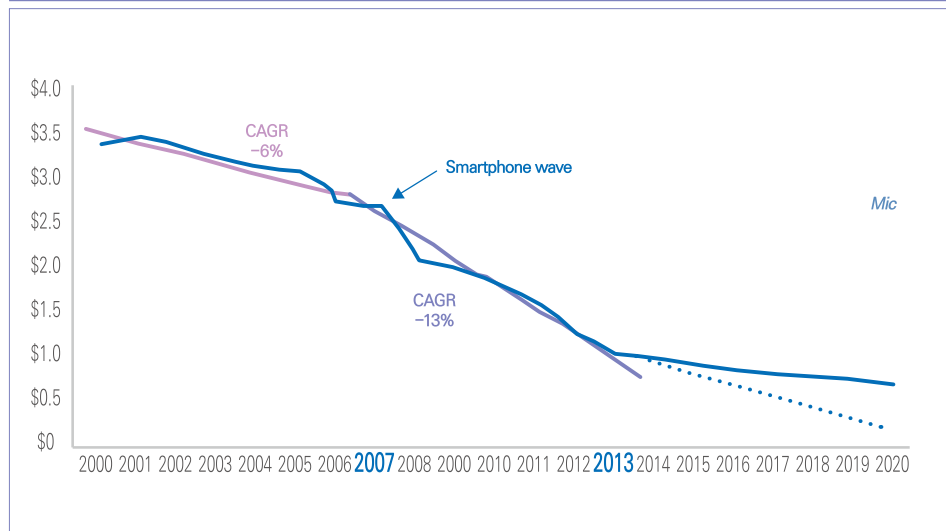
먼저 전통 제조업, 빌딩 및 도시 인프라, 농업 등에서 IoT 센서의 수요가 증가하고 있다. 불보건설기계는 산업용 차량에 자율 운행이 가능하도록 센서를 장착하여 배기가스 저감 및 인명 사고를 감소시키는 효과를 기대하고 있으며, 제조사와 파트너들이 스마트폰 앱을 통해 부품 교체 시점을 알 수 있도록 하고 있다. 라스베이거스에서는 IoT 장비를 통해 데이터를 전송받아 교통 문제와 도로 파손을 모니터링하는 데 사용하고 있다. 또한, 린드세이는 농업용 기계에 센서를 부착해 필요한 물의 양, 영양소 수준, 토양 상태, 하이퍼 타겟 농지 면적의 특성 등 다양한 데이터를 일기예보 등과 연계시켜 농부가 작물을 관개하는 데 이용하도록 하고 있다.

또한 IoT 센서를 제조함에 있어, 기능은 다르나 제조 공정이 유사한 센서는 통합하는 추세이다. 이를 통해 '다품종 소량 비즈니스'라는 한계를 완화하고 있다. 독일의 보쉬社は 압력과 온도를 동시에 측정하는 환경 콤보센서(BMP280)를 개발한 이후, 압력·온도·습도·가스 기능이 통합되어 향상된 BME680 센서를 연이어 출시했다.

IoT 센서 HW가 점차 범용화되면서 센서 성능이 주요 차별적 요소가 되었으며, 맞춤형 SW 애플리케이션이 주목받고 있다. GE는 스마트 팩토리에 자사 솔루션 '디지털 파워 플랜트'를 제공함에 따라 발전소 가동 시간 2.5배 향상했다.

IoT 센서를 설계함에 있어 그 주도권이 센서솔루션업체로 이동하고 있다. 스마트 센서 관련 하드웨어는 일부 대기업에 의해 과점될 것으로 전망되며, 주문형 센서는 시스템 반도체 시장과 같이 센서 설계업체와 위탁생산업체가 공존하는 생태계가 조성될 것으로 전망된다. 또한, 센서 범용화에 따라, 센서 가격은 하락추세를 보이고 있다. 이로 인해 하드웨어 업체의 수익구조는 다소 악화되고 있다.

〈그림 III-7〉 MEMS 센서의 평균 단가



출처: 포스코경영연구원(2018), Yole development, 2018년 1월 기준

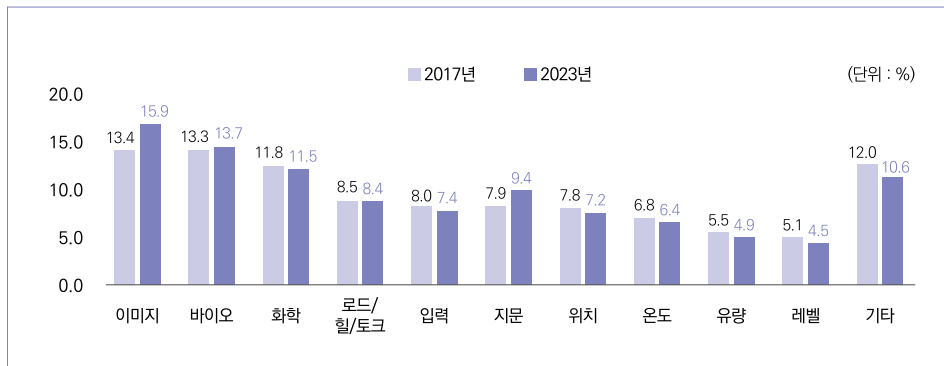
스마트 센서 분야의 산업가치는 기존 센서의 하드웨어에서 솔루션을 보유하거나, 시스템 설계가 가능한 소프트웨어 업체로 이동하고 있다. 센서 관련 하드웨어 업체 위상이 저하되는 것에 비해, 최근 다양한 이종 기기 등장에 따라, 이를 통합하고 전환시키는 소프트웨어의 중요성이 높아지고 있다. 통합 플랫폼 운영사가 스마트폰, 자율주행차 등에 적용할 센서를 결정할 수 있으므로, Google, Apple 등 O/S 설계 기업이 업계를 주도할 것으로 전망된다. 스마트 센서는 기술 주도형에서 응용 주도형으로 전환되고 있으므로, 솔루션이 필수 기술로 대두되고 있다.



#### 다. 주요 IoT 센서 구분

IoT 센서 주요 품목별 시장 규모는 이미지, 바이오, 화학 센서가 가장 크며, 이러한 3대 제품이 센서 산업의 약 40%를 차지한다. 이미지 센서는 스마트폰 탑재량 증가, 타 센서 대비 높은 가격 등으로 인해 시장 규모가 가장 크며 2017년 기준 센서 산업에서 차지하는 비중은 13.4%로 나타났다. 바이오 센서는 전체 센서 산업의 13.3%를 차지하며 혈당 측정 센서가 성장을 견인해오다 최근 헬스케어 패러다임이 치료에서 사전 예방으로 전환되며 더욱 큰 성장 가능성을 보여주고 있다. 화학 센서는 전체 센서 산업의 11.8%를 차지하며 산소, 이산화탄소 등의 농도 등을 감지하여 가스 누출을 예방하는 활동 등에 활용된다.

〈그림 III-8〉 종류별 IoT 센서의 점유율



출처: BCC 리서치(2017), 반도체산업협회

#### 라. IoT 센서와 활용

IoT 센서는 특정 환경 변화를 측정하는 수준에서 나아가, 데이터 처리, 자가 진단 등이 가능한 스마트 센서 기술로 전환이 가속화되고 있다. 과거의 센서는 특정 물리/화학/기계적 변화를 측정하는 수준에 머물렀으며 센서의 수가 증가함에 따라 대용량 데이터를 중앙 집중 방식으로 처리하기에는 전력 소모가 커지는 단점이 존재했다. 하지만, MEMS 및 SoC(System on Chip) 기술의 발전과 MCU가 내장되면서 데이터 처리, 저장, 자동 보정, 자가 진단, 의사 결정, 통신 등의 기능을 수행하는 스마트 센서 중심으로 발전하고 있다.

〈그림 III-9〉 스마트 센서 주요 구조



출처: ETRI(2019), 스마트 IT융합 플랫폼을 위한 지능형 센서 기술 동향

스마트 센서의 주요 구조는 특정 환경 변화를 감지하는 센서, 데이터를 수집·분석·처리하는 MCU, 데이터를 서버로 전송 또는 수신받는 통신 모듈이 하나의 반도체로 집적화되어 있다. 센서 자체에서 데이터를 가공·처리하여 서버로 전송하는 분산 방식이 가능하며, 중앙 처리 장치의 전력 소모를 줄이고 유연한 생산 방식이 가능하며 제조업 스마트화를 위한 핵심 기술로 부상하고 있다. 센서를 통해 생성된 데이터는 실질적인 자동화·스마트화의 기초가 되며, 상태모니터링 및 진단을 통해 설비 등의 예지와 보전 등에 활용되고 있다.

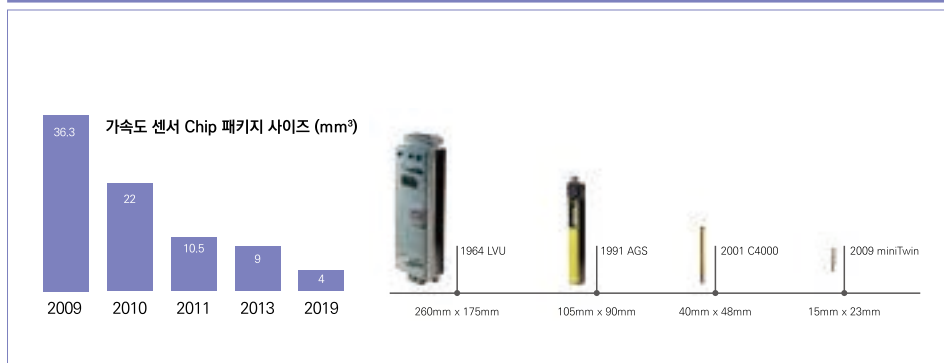
〈표 III-4〉 IoT 센서의 발전

구분	시기	특성
1세대 (Discrete Sensor)	1970~1980	• 온도, 압력, 가속도, 변위 등의 물리량을 전기적 신호로 변환하는 기능의 센싱소자와 증폭, 보정, 보상의 신호처리 회로가 별개로 분리
2세대 (Intergated Sensor)	1980~1990	• 센서의 잡음 성능을 높이고 소형화하기 위해 센서와 신호처리 회로가 결합된 형태로 제작, MEMS 기술이 도입
3세대 (Digital Sensor)	2000년대	• CMOS 기술의 발전으로 아날로그 회로에 디지털 회로가 집적되면서 센서의 이득, 오프셋, 비선형 등을 디지털 방식으로 보정하고, 보정 데이터를 비휘발성 메모리에 저장
4세대 (Smart Sensor)	2012 이후	• MCU가 센서에 내장되고 SoC 기술이 접목, MCU의 제어, 판단, 저장, 통신 등의 기능을 활용하여 센서의 성능 향상과 다중센서, 네트워크 센서, IoT 센서로 진화

또한, 제조 공정에 사용되는 센서는 무선화·지능화로 발전됨에 따라 소형화, 복합화, 저전력 중심으로 진화가 가속화되고 있다.

**(소형·경량화)** 제조 장비에 적용되는 센서는 크기는 작고 높은 성능을 제공하는 센서 기술이 요구된다. 하나의 칩에 센서와 반도체 회로를 집적한 MEMS(Micro Electro Mechanical System) 기술과 SoC 기술이 진화됨에 따라 고성능·소형화 기술 개발이 활발히 진행되고 있다. 이렇게 소형화된 IoT 센서는 설치 장소의 제약에서 자유롭고, 밸브, 액추에이터 등 소형 부품에 통합 가능하며, 작은 크기로 최상의 성능을 제공하므로 다양한 애플리케이션에 적용이 가능하다. 일례로 가속도 센서의 칩 크기는 2009년에 비해 1/9 수준으로 크기가 줄어 초소형으로 개발 중이며, 이에 따라 제조 장비도 점차 소형화되는 추세이다.

〈그림 III-10〉 센서의 소형화와 산업용 장비의 진화



출처: Maxim Integrated(2018)

**(복합 센서)** 복수의 센서를 단일 모듈로 집적화하여 서로 다른 감지 데이터를 결합하여 불확실성을 낮춰 더욱 정확한 정보를 생성하는 복합 센서 모듈이 개발되는 추세이다. 보쉬社は 압력과 온도를 동시에 측정하는 복합 센서를 개발한 이후 다양한 환경 측정용 센서를 통합하여 출시했다. ST마이크로일렉트로닉스社は 가속도 센서, 압력 센서, 습도 센서가 결합된 멀티 센서 기반 제조용 스마트 센서를 출시했다.

〈그림 Ⅲ-11〉 제조 산업용 복합 센서 제품

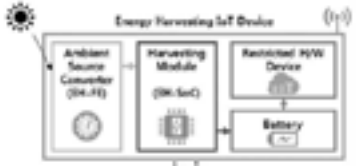
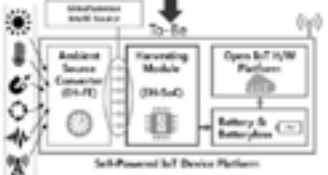


출처: 보쉬, ST마이크로일렉트로닉스

**(저전력 센서)** 제조 공정에서 무선 센서 기술 활용이 가속화되면서 장시간 유지가 가능한 전력이 요구되고 있다. 이를 위해 적은 전력으로도 센서를 가동하는 저전력 기술 개발이 활발하다. 제조 공정과 같이 지속적인 전력 공급이 필요한 센서를 위해 전력 배선(유선)을 사용하거나, 또는 무선 센서의 경우 배터리를 이용하고 있다. 이는 배터리 교체 등 지속적인 유지 보수와 설비 재배치 등에 시간과 비용이 증가하는 문제가 있다. 때문에 전력 소비를 줄이기 위해 대부분의 센서들은 데이터를 송수신할 때 주로 동작한다. MIT 연구팀은 대기 상태에서 센서가 소비하는 전력을 최대 100배까지 줄일 수 있는 송수신기용 회로를 설계하여 센서의 배터리 수명을 수개월까지 확대할 수 있다고 밝히기도 했다.

또한, 에너지 자립을 통해 배터리 교환과 전기 배선의 유지관리 비용을 줄일 수 있는 MEMs 기반의 에너지 하베스팅 기술이 각광받고 있다. 에너지 하베스팅 기술이란 빛, 진동, 압력, 열, RF(Radio Frequency) 등 일상생활에서 버려지거나 소모되는 에너지를 수집하여 전력으로 변환해 사용하는 기술을 말한다. 다양한 환경에서 요구하는 전력을 충족하기 위해 단일 에너지보다는 다중 에너지 수집이 가능한 에너지 하베스팅 기능을 IoT 플랫폼에 연동하고자 하는 연구가 활발하다.

〈표 III-5〉 에너지 수집 기술

As-is (단일 에너지 수집)	To-be (다중 에너지 수집)
	
<ul style="list-style-type: none"> <li>제한된 환경에서 단일 입력 에너지만 수집 가능한 하베스터(제한된 전력 수집)</li> <li>사용 용도가 한정된 주문 생산형 IoT 디바이스로 제한</li> </ul>	<ul style="list-style-type: none"> <li>다양한 환경에서 다중 입력 에너지 수집 및 센싱이 가능한 하베스터(응용에 따라 전력 수집 가변)</li> <li>사용 용도 변경이 용이한 범용성 IoT 디바이스 연동 가능</li> </ul>

### 1-2 IoT 연결 기술 활용과 설계

무선통신은 통신 거리에 따라 근거리 무선통신과 원거리 무선통신이 있으며, 많은 IoT 솔루션의 경우 저전력·저지연의 요구 사항을 충족하기 위한 기술 개발이 활발하다.

무선통신의 주요 특성으로 외부 전원을 연결하기 어렵고, 내장 배터리 용량의 한계 등 제약으로 인해 유지·보수 비용을 절감하기 위해 전력 소모 최소화를 위한 저전력 네트워크 기술이 요구된다. 또한, IoT 디바이스가 점차 확대됨에 따라 전송되는 데이터의 양과 속도 등에 따라 전송 시간이 지연될 수 있기에 송수신되는 데이터 트래픽을 최소화하기 위한 저지연 네트워크 기술이 요구되고 있다.

**(근거리 무선통신)** 근거리 무선통신의 경우 Wi-Fi, WirelessHART(이하 WHART), ISA100, Bluetooth 등의 근거리 통신이 적용되고 있다.

〈표 III-6〉 근거리 무선통신기술 비교

구분	WiFi	NFC(RFID)	Bluetooth	ZigBee
주파수 대역 (비면허대역)	2.4GHz 5GHz 60GHz(초고속 통신용)	13.56MHz	2.4GHz	2.4GHz
전송거리	약 100m	10cm 이내	1~100m	100m
전송속도	802.11 b/g/n~300Mbps 802.11 ac~1.3Gbps	424kbps	2Mbps	250Kbps
응용분야	인터넷 무선 LAN	건물출입 인증, 결제, 교통카드	wearable(스마트 워치, 만보기, 마우스, Music player) 소형 매장O2O 서비스 (대형매장은 다수 비컨필요)	스마트 홈
소비전력	평균 100mW	X	BLE 5mA	30mA
특징	전력소모 큼, Coverage 크다	무전원 동작, 보안 우수	저전력 가능, Coverage 작다	저전력, 저비용, 별도의 브리지 장치 필요

**(원거리 무선통신)** Wi-Fi, Bluetooth 등 근거리 무선통신기술의 단점으로 통신 거리가 있으며, IoT 전용 저전력 원거리 무선통신기술인 LPWA(Low Power Wide Area)의 등장으로 활용이 확대되고 있다.

근거리 무선통신기술은 10M~200M 이내의 근거리에서 주로 사용되나, 원거리 통신이 필요한 상황에서는 통신 거리를 만족하기에는 한계가 있다. LPWA 기술은 사물인터넷 디바이스들을 위한 이동통신망으로 10년 이상의 긴 배터리 수명과, 수십 km의 도달거리로 넓은 범위와 설치비용의 장점을 가지고 있다. 대표적인 기술로는 크게 SigFox와 LoRa, LTE-M과 NB-IoT 등을 꼽을 수 있으며, 각 기술별 장단점이 존재하기 때문에 제조 산업에 적용하기 위한 기술 경쟁이 활발하다.

LTE-M, NB-IoT는 기존 무선전화 통신망의 대역폭을 줄여 장거리 저전력 통신 네트워크를 구현하는 방식이다. 기존의 인프라를 사용하기 때문에 초기 투자 비용이 낮은 반면, 통신 사용료를 지급해야 하고, 통신사의 네트워크를 거쳐야 하기에 데이터 보안 문제 등의 단점이 있다. 해외의 경우 LTE-M은 AT&T, Verizon 등 미국 통신사 중심으로 이용되고 있으며, NB-IoT는 Hwawei, Intel, Qualcomm 등 제조사와 Ericsson, Nokia, KT, LGU+ 등 통신사 기업이 이용하고 있다.

LoRa, SIGFOX 등의 경우에는 사설 Gateway 설치 등 초기 망 구축 비용이 낮으며, 통신 방식을 사용자화 할 수 있어 데이터 보안성을 높일 수 있다는 장점이 있다. LoRa는 LoRa Alliance인 KPN, Swisscom, ZTE, SKT 등이 있으며, SigFoX는 유럽 기업 중심으로 상용화되고 있다.

〈표 III-7〉 저전력 광역 무선통신기술 비교

구분	LTE-M	NB-IoT	LoRa	SigFox
통신 거리	~11Km	~15Km	~11Km	~13Km
통신 속도	~1Mbps	~150kbps	~10kbps	~100bps
주파수 대역	면허 대역 (1.4MHz)	면허 대역 (200KHz)	비면허 대역 (900MHz)	비면허 대역 (900MHz)
장점	- 전국 서비스 가능 - 통신 품질의 안정성 - 기존 네트워크 활용	- 통신 품질의 안정성 - 실내 커버리지 가능	- 저전력 장거리 통신 - LTE-M 모듈의 1/4 가격 - 저렴한 구축 비용	- 저전력 장거리 통신
단점	- 고가의 통신 모듈 가격	- LoRa에 비해 통신 모듈 가격이 비쌌	- 비면허 대역으로 네트워크 불안정	- 전세계 한정적 인프라 - 비면허 대역으로 네트워크 불안정

출처: ETRI(2019), “스마트 IT 융합 플랫폼을 위한 지능형 센서 기술 동향”

**(5G 통신)** 센서 및 디바이스의 수가 증가함에 따라 데이터의 양이 급증하나 4G의 전송지연(20~60ms)과 Wi-Fi의 QoS 문제(간헐적 끊김, 전파간섭) 등으로 고신뢰성과 저지연이 요구되는 환경에서 활용하는 데 한계가 있다. 또한, IoT 전용망의 경우 저렴한 가격으로 망 서비스를 제공할 수 있는 장점은 존재하나, 활용하기 위한 새로운 전용 하드웨어 기기 및 별도의 전용 서비스 등을 개발, 구축, 운영해야 하므로 많은 비용이 발생하고 대용량 IoT Connectivity를 수용하는 데 한계가 있다.

5G 기술은 차세대 저비용 IoT 네트워크로 각광받고 있으며, 초저지연, 초고속, 초연결, 초저전력의 네 가지 특성은 기존 통신기술에 비해 확장성 및 유연성을 증가시킬 것으로 기대된다.

- **초저지연:** 무선으로 대용량 데이터 전송과 1ms 이하의 저지연 통신이 가능하다.
- **초연결:** 1km<sup>2</sup> 영역 안에서 기존 4G 대비 10배인 100만 개 이상의 장치가 동시에 통신·연결이 가능하다.
- **초고속:** 5G 기술은 기존 LTE 대비 20배 빠르며 데이터 처리 용량도 100배 많아 실시간 고용량 데이터 전송, 로봇의 자율이동을 통한 예측 불가능한 생산 공정 문제를 해결할 수 있다.
- **저전력:** 모든 디바이스의 배터리 수명도 기존 LTE 대비 10배 이상 증가한다.



〈표 III-8〉 5G와 주요 유무선통신기술과의 성능 비교

특성	유선(TSN)	4G(LTE-adv.)	Wi-Fi(802.11ax)	5G
이동성	×	○	○	○
전송지연	1ms	20~60ms	2ms~200ms	1ms
QoS보장	○	○	×	○
최고전송속도	10Gbps	1Gbps	1.2Gbps	20Gbps
최대연결기기	-	100,00m/km2	-	1,000,000/km2
에너지효율	-	1×		100×
설치-유지보수	복잡	단순	단순	단순

출처: 과학기술정보통신부(2019), 5G기반스마트공장고도화전략(안)

## 2 IoT 도입 사례

‘스마트지킴이2’은 2021년 SK텔레콤이 출시한 IoT 기반 웨어러블 위치확인 장비이다. 이 웨어러블 기기와 전용 앱을 통해 치매를 겪는 노인이나 발달장애인, 미취학 아동 등 취약계층의 안전한 돌봄이 가능하다. ‘스마트지킴이2’는 2019년 출시된 스마트지킴이의 후속 제품이다. 특히 기존 제품에 비해 IoT 기반 위치 확인 기능이 강화되었다. 이를 위해 전국을 커버하는 사물인터넷 전용망인 LTE Cat.M1과 GPS, 무선랜을 활용하는 ‘복합 측위 시스템’을 적용해 실내·외 제약 없는 위치 확인이 가능하게 했다.

〈그림 III-12〉 SK텔레콤 ‘스마트지킴이2’



출처: SK텔레콤

보호자는 전용 앱에서 착용자의 활동 권역을 설정할 수 있으며, 착용자가 설정된 권역을 벗어날 경우 이에 대한 알림을 받는 동시에 필요시 여러 명에게 착용자 위치 정보를 공유할 수도 있다. 또한 보호자는 착용자의 상황이 좋지 않다고 판단할 경우 “걱정됩니다. 전화주세요” 등의 간단한 긴급 메시지를 웨어러블 기기로 보낼 수 있으며, 착용자는 기기의 SOS 응급호출 버튼을 눌러 쉽고 빠르게 보호자에게 도움을 요청할 수 있다. 또한 보호자가 착용자의 위치를 확인할 때 측위 방식을 고려한 오차 범위 정보를 제공해 착용자가 실종시 수색 범위를 가늠할 수 있게 했다.

이 밖에도 ‘스마트지킴이2’는 다양한 센서를 포함하여 착용자의 심박수나 산소포화도, 활동량(걸음 수), 수면건강상태(수면혈중 산소포화도) 등의 건강정보와 착용자의 낙상감지 기능도 탑재하고 있다. ‘스마트지킴이2’는 IP68등급의 방진 방수 및 과열 감지나 자동차단 기능을 지원하며, 1회 배터리 충전으로 2~3일 정도 사용할 수 있다.

구글 산하 네스트(Nest)는 손바닥 크기의 둥근 형태를 가진 온도조절장치 ‘네스트 학습형 온도조절기(Nest Learning Thermostat)’가 주력제품이다. 네스트 온도조절기는 일반 온도조절기와는 달리 스스로 사용 패턴을 학습하여 맞춤형 실내 환경 제어 서비스를 제공한다.

〈그림 III-13〉 네스트 학습형 온도조절기



출처: 구글 네스트

네스트 온도조절기는 일주일 동안 사용자가 온도를 설정하는 패턴을 학습하고, 그 후에는 지능적으로 냉·난방을 제어하여 수동 조작 없이 자동으로 적당 온도를 조절해주는 인공지능 냉난방 조절장치이다. 또한, 집안 복사열과 외부 날씨에 따라 난방과 냉방의 목표 온도까지 도달하는 시간을 표시해 주며, 이 데이터를 고려하여 최대한 효율적으로 동작한다. 이를 통하여 집안에 사람이 있는 시간(퇴근, 하교 후 ~ 출근, 등교 전) 동안 다수의 복사열과 냉난방 장치를 활용하여 적절한 온도를 유지하다가 사람이 없는 시간(출근, 등교 후 ~ 퇴근, 하교 전) 동안은 에너지를 최대한 절전하게 된다. 또한 동작 감지 센서가 있어서 집안에 사람의 움직임이 없으면 자동으로 절전모드로 전환하는 기능을 가지고 있다. 사용자가 접근하면 이 센서가 감지해 디스플레이가 켜지고 현재 온도, 시간 등 필요 정보를 시각적으로 디스플레이해 불필요한 전기 낭비를 줄인다.

특히 본 기기는 난방과 냉방을 집 안에서 사람의 지시를 받아 조절하는 것뿐만 아니라 최근 유행하는 스마트홈 기능을 일찍이 도입하여 사람이 집 외부에서도 온도를 통제할 수 있고, 또한 날씨 데이터를 수신하여 이를 온도 조절에 사용하도록 하기 위하여 IoT 기술을 통한 인터넷 망 접속이 가능하도록 하였다. 이를 통해 네스트 온도조절기는 최소한의 에너지 소모로 적절한 내부 온도를 유지하고, 필요시에는 시스템 업그레이드도 편하게 진행할 수 있다. 또한 에어컨의 실외기가 중단되어도 팬은 계속 돌면서 냉각 증발기 코일을 통해 일정 시간 동안 시원함을 유지할 수 있는데, 네스트 온도조절기는 이러한 특성을 이해해서

에어컨의 동작시간을 30% 절약할 수도 있다. 이러한 지능형 IoT 서비스를 위해 네스트 온도조절기는 머신러닝과 딥러닝 소프트웨어를 탑재하고 있는 것으로 알려져 있다. 네스트는 온도조절기 이외에도 네스트 허브, 네스트 캠, 네스트 도어벨 등 다양한 스마트홈 IoT 제품들을 판매하고 있다.

바이탈리티(Vitality)의 스마트 약병 '글로캡(GlowCap)'은 약을 먹는 시간을 알려주는 약병이다. 처방받은 약을 의사가 알려준 대로 제 때 그리고 정확한 양으로 복용하는 것은 치료의 기본이다. 하지만, 특히 노년층의 경우 때때로 이를 제대로 지키지 못하는 경우가 있다. 글로캡은 이러한 문제점을 해결하기 위하여 환자들이 제시간에 약을 복용할 수 있도록 도와주는 스마트 약병이다.

〈그림 III-14〉 바이탈리티의 글로캡



출처: NantHealth

글로캡의 주요 두 가지 기능으로는 환자의 약 복용 시간을 알려주는 것과 의사에게 복용 결과를 전달하는 시스템이다. 먼저 복용 시간이 되면 약병 뚜껑의 램프가 켜지면서 동시에 알람 소리로서 환자에게 이를 알린다. 만일 이러한 알람에도 불구하고 환자가 약병을 열지 않으면, 환자에게 추가적으로 전화를 걸어 복용이 필요함을 상기시킨다. 이처럼 다양한 알람 기능이 가능한 까닭은 글로캡이 AT&T와 같은 무선망으로 바이탈리티의 네트워크와 연결되어 있기 때문이다. 따라서 사용자가 약을 먹기 위해 약병을 열면 이를 센서가 감지하고 통신망을

통해 약을 먹었다는 정보가 바이탈리티 서버로 전송된다.

사용자 관찰 결과 글로캡 사용자들의 복약률은 98%까지 높아졌다고 한다. 사용자의 복약 여부는 바이탈리티의 보안 네트워크로 전송되어 저장되고, 이렇게 매일 차곡차곡 쌓인 복약 내역은 주간, 월간 단위로 정리되어 환자, 보호자, 주치의에게 제공된다. 이를 통해 환자의 담당 의사는 만약 환자가 진료를 받아왔더라도 계속 같은 증상을 보인다면 그 원인이 약을 제때 복용하지 않아서인지 혹은 다른 약을 처방해야 하는지 등의 판단을 할 수 있다.

일본의 소프트웨어 전문기업 옵팀(OPTiM)은 드론 등을 이용한 핀포인트 농약살포 기술을 개발하였다. 일반적으로 농약 살포 작업은 농가에 부담이 크다. 희석한 농약을 담은 탱크를 메거나 차에 싣고 농장 안으로 들어가 살포하는 작업은 매우 고된 작업이다. 더군다나 차츰 고령화되어가는 농가에게는 이러한 농약 살포 드론 등이 매우 효율적일 수 있다.

옵팀의 핀포인트 농약살포 기술은 검출된 병충해가 있는 위치에만 농약을 살포하여, 농약을 살포하기 위한 노동력을 감소시키고, 사용하는 농약의 양과 비용을 절감하며, 이를 통해 자연환경과 농작물에 미치는 악영향을 최소한으로 줄인다. 옵팀은 비용 때문에 솔루션 도입을 망설이는 농가를 대상으로 우선 IoT 제품을 무상 보급했다. 이어서 해당 솔루션으로 생산된 농산물을 전량 매수해 생산자와 수익을 분배하는 새로운 비즈니스 모델을 구축했다. IoT 솔루션 농가의 경우, 농약 사용량을 90%나 줄인 저농약 브랜드 농작물로 특화해 수익이 증가한 성과도 나타났다.

〈그림 III-15〉 옵팀 핀포인트 농약살포 기술



출처: 옵팀 유튜브

### 3 IoT의 보안

사물인터넷이 B2C, B2B 시장의 스마트홈, 스마트워치, 스마트팩토리 등으로 점차 확대보급되면서 다양한 보안 위협문제가 대두되고 있다. 특히 우리 주변에서 쉽게 볼 수 있는 가전, 웨어러블 기기, 각종 스마트팩토리 센서, 자동차 등의 사물도 보안위협 대상이 되고 있다. 이 중 자동차 해킹, 교통 제어 시스템 해킹, 전력 시스템 해킹 등과 같은 사물인터넷 상의 보안 사고는 심각한 인명피해 등의 손실을 미칠 수 있다.

〈표 III-9〉 사물인터넷 공격 및 위협사례

유형	공격 및 위협사례
악성코드 감염	2013년 말부터 2014년 초 Proofpoint는 10만 대 이상의 스마트 TV, 스마트 냉장고, 라우터 등의 Thingbots을 통해 총 750,000건의 피싱과 스팸 메일을 전 세계 개인과 기업에 발송 되었음을 밝힘
	리눅스 달로스 웹으로 PHP 취약점을 악용하여 보안용 IP 카메라, 셋톱박스, 무선랜카드 등 리눅스 OS를 사용하는 사물인터넷 기기 감염
데이터 및 기기 탈취	자동차 해킹하여 디지털 콤팩스, 휠 인코더, 관성 측정 유닛 등의 센서에 잘못된 정보를 삽입함으로써 급정거하거나 차선을 이탈 등 조작 시연(2013 블랙햇 컨퍼런스)
펌웨어, 운영체제	Hack in the Box 보안컨퍼런스에서 무선 IP카메라의 펌웨어 변형공격 시연
권한 탈취, 불법 접근 및 정보유출	2012년 미국 데이터 암호화 및 인증 정착의 부재로 고속도로 교통표시판(VMS) 및 교통제어 시스템 해킹
	2012년 OBD-II (자기진단장치)의 취약점을 이용 차량 접속을 통해 BMW 차량 300대 이상 도난
	2013년 독일 리큐리티랩스(Recurity Labs) 해킹 실험을 통해 에틀링겐(Ettlingen) 도시의 전력 공급을 외부에서 무단으로 차단할 수 있음을 입증
	SHODAN은 필요 없는 사물인터넷 디바이스와 서비스들의 정보를 매달 5억개 이상 수집하여 검색 기능을 제공해주는 서비스(2009년 서비스 시작, 백도어를 검색하기 때문에 어둠의 구글이라고 불림)로써 SHODAN을 이용하여 취약한 사물을 검색하면 디폴트 패스워드를 사용하고 있는 프린터, 서버, 시스템제어 장치들이 검색되어 아무런 인증절차 없이 사물인터넷 기기에 접속 가능

출처: 사물인터넷 시대의 사이버 물리 시스템 보안 기술 동향(2015)

사물인터넷 환경에서는 다양한 사물과 이들에 장착된 센서들이 인터넷 등을 통해 서로 연결됨에 따라 우리가 흔히 생각하는 기존 인터넷 보안위협이 사물인터넷에서도 발생할 수 있다. 나아가 사물인터넷 환경에서의 사물들은 필요에 따라 소형화 및 경량화된 구조에서 동작하거나 저전력 통신 기술을 이용한 네트워크에서 작동하기 때문에 기존의 환경에 비하여 보안기술을 사용하는데 많은 제한사항이 있다. 따라서 위의 표에서 나타나는 악성코드 감염, 데이터 및 기기탈취, 펌웨어(firmware) 공격, 불법 접근 및 정보유출 등의 보안위협이 발생할 수 있다.

사물인터넷 환경에서 제품과 서비스를 제공하는 기업들은 이와 같은 보안위협에 대비하고 대응하기 위해서 다양한 보안상 고려사항을 준수해야 한다.

### 3-1 센서 보안

사물인터넷 '센서' 분야와 관련된 보안사항의 종류는 다양하다. 이제까지의 보안위협 사례는 하드웨어를 포함한 다른 영역보다는 센서 내에 내장된 펌웨어와 같은 소프트웨어를 손상시켜 보안을 침해하는 경우가 많았다.

#### 가. 보안 부팅

컴퓨터의 전원이 켜지고 첫 번째로 실행하는 소프트웨어를 '부팅'이라고 한다. 보안 부팅은 디지털 서명(Digital Signature)이 없는 소프트웨어의 실행을 원천적으로 방지하여 전체 부팅 프로세스를 보호한다. 보안 부팅이 시작되면 펌웨어에 키(key)값을 입력할 수 있다. 일단 키값이 입력되면 보안 부팅은 키값이 있는 소프트웨어만 로딩시켜 기타 위험요소를 가진 소프트웨어의 실행을 막을 수 있다.

#### 나. 악성 소프트웨어

우리의 노트북이나 PC에는 대부분 바이러스 백신 소프트웨어를 탑재하고 있다. 바이러스 백신 소프트웨어는 컴퓨터를 손상시킬 수 있는 악성 소프트웨어를 감지하기 위해 만들어졌다. 악성 소프트웨어에 의한 피해 사례로 이란의 우라늄 농축 시설의 원심 분리기를 해킹한 스텍스넷(Stuxnet) 바이러스가 있다. 스텍스넷은 단순히 PC를 감염시키는

것이 아니라 기계를 제어하는 PLC(Programmable Logic Controller)를 타깃으로 하여 물리적인 피해를 입힌다. 이란의 사례에서는 그 타깃이 농축 우라늄을 생산하는 기계인 원심 분리기였기에 그 심각성이 컸다. 스텍스넷은 MS 윈도 OS 및 네트워크를 사용하는 기계들을 타깃으로 해킹 공격을 한 후 그 중 Siemens Step 7 소프트웨어를 찾아 감염시켰다. 스텍스넷은 이란 핵 원심 분리기의 약 1/5을 파괴했다고 전해진다.

#### 다. 양성 소프트웨어

패치는 컴퓨터 소프트웨어를 업데이트하도록 설계된 소프트웨어의 한 종류이다. 일부 패치는 기존 소프트웨어의 성능을 향상시키도록 설계되기도 하지만, 보안 취약점을 수정하기 위해 고안된 패치도 있다. 패치를 적용하고 관리하기 위해서는 어떠한 지정된 시간에 어떤 시스템에 어떤 패치를 적용해야 하는지를 전략적으로 계획해야 한다. VMware, 오라클, 마이크로소프트 같은 회사는 보안 관련 패치를 매년 수백 개씩 내놓는다.

### 3-2 연결 네트워크 보안

#### 가. 네트워크 보안

네트워크 상의 보안은 일반적으로 사용자 아이디와 암호를 사용하여 인증하는 것으로 시작된다. 이를 단일 요소 인증(One-factor Authentication)이라고 한다. 그러나 보안에 대한 관심이 높아지면서 많은 업체가 이중 요소 인증(Two-factor Authentication)을 채택하고 이를 구현하고 있다. 이는 모바일폰, OTP 등 물리적으로 소유하고 있는 장비 등을 사용하여 인증하는 방법이다.

일단 인증되면 방화벽(Firewall)은 네트워크 사용자가 액세스할 수 있는 서비스를 규정한다. 방화벽은 정해진 보안 규칙에 따라 수신 및 발신 네트워크 트래픽을 모니터링하고 제어한다. 이러한 보안 규칙은 일반적으로 신뢰할 수 있는 내부 보안



네트워크와 보안성과 신뢰성이 낮은 것으로 간주되는 다른 외부 네트워크(인터넷 등) 사이에 장벽을 설정한다. 이를 통해 신뢰할 수 없는 네트워크로부터 신뢰할 수 있는 네트워크를 보호한다.

또한, 방화벽은 네트워크를 통해 전염되는 컴퓨터 웜이나 트로이 목마와 같은 소프트웨어 바이러스를 방지한다. 백신 소프트웨어 또는 침입 방지 시스템은 이러한 불량 소프트웨어를 탐지하거나 차단하는 것을 도와준다.

마지막으로 방화벽은 연결 계층에서 전송되는 암호화된 데이터를 보호한다. 이를 위해 방화벽은 송신자와 수신자가 공유하는 키값들을 사용하여 전송된 데이터를 열람한다. 이에 암호화된 데이터의 전송 방법 뿐만 아니라, 키값에 대한 접근과 제어 역시 보안에 있어서 중요해졌다. 이렇듯 보안 위협이 증가하고 사물과 컴퓨터가 점점 더 네트워크화 되어 원거리에서도 액세스할 수 있게 됨에 따라 보안 분야에서는 많은 혁신이 계속되고 있다.

### 3-3 사물인터넷 보안 대응 방안

IoT 보안 위협에 대해서도 다양한 대응 방안이 존재한다. 사물인터넷도 기본적으로는 네트워크를 사용하기 때문에 기밀성(confidentiality), 무결성(integrity), 가용성(availability) 세 가지 요소를 보안의 기본 요소로 갖추어야 한다. 이에 IoT를 활용하여 서비스를 제공하고자 하는 회사들은 기존과 동일한 보안수준을 유지하기 위하여 기 활용하던 보안 대응 방안을 사물인터넷 환경에도 동일하게 적용할 필요가 있다. 아래 표에는 OWASP에서 공개한 IoT Top 10 취약점 및 대응방안을 정리한 것으로 대부분 기존과 크게 다르지 않음을 알 수 있다.

〈표 III-10〉 OWASP IoT Top 10 취약점

ID	취약점 명	설명	대응방안
I1	취약한 웹 인터페이스 (Insecure Web Interface)	취약한 계정관리, SQL삽입, XSS 등 웹애플리케이션 취약점 존재	디폴트 계정 사용금지, 암호찾기 절차 강화, SQL삽입 등 웹애플리케이션 취약점 제거
I2	불충분한 인증/권한 (Insufficient Authentication Authorization)	단순한 패스워드 사용, 쿠키재전송, 복호화 가능한 알고리즘을 이용한 패스워드 암호화 등을 통한 인증/권한 획득	패스워드 복잡도 강화, 단방향 암호알고리즘을 이용한 패스워드 암호화, 2팩터 인증 등 사용
I3	취약한 네트워크 서비스 (Insecure Network Services)	불필요한 서비스 포트 사용, 취약한 UDP 서비스 사용 등 네트워크 취약점 존재	필요한 포트만 접속을 허용하는 등 네트워크 취약점 제거
I4	전송암호화/ 무결성검증 취약 (Lack of Transport Encryption/ Integrity Verification)	중요정보의 평문전송, SSL/TLS의 잘못된 보안설정 등으로 인한 전송구간 중요정보 노출	통신구간은 안전한 보안설정이 이루어진 SSL/TLS 등을 이용하여 기밀성 보장
I5	프라이버시 문제 (Privacy Concerns)	I 1~ I 4 취약점을 이용하여 개인정보 유출	불필요한 개인정보는 네트워크 구간에 전송하지 않고, 전송 시 반드시 암호화
I6	취약한 클라우드 인터페이스 (Insecure Cloud Interface)	I 1~ I 4 취약점을 이용하여 클라우드 계정정보 탈취 또는 중요정보 유출	I 1~ I 4 대응방안 준수
I7	취약한 모바일 인터페이스 (Insecure Mobile Interface)	I 1~ I 4 취약점을 이용하여 모바일앱 계정정보 탈취 또는 중요정보 유출	I 1~ I 4 대응방안 준수
I8	불충분한 보안 설정 가능성 (Insufficient Security Configurability)	사물인터넷 관련 서버, 네트워크 장비, 기기 등의 보안설정이 미흡하여 계정관리 취약, 보안감사 및 로깅 미실시	관리자와 일반사용자 계정을 분리 사용하고 강력한 계정관리 보안설정 준수, 보안감사 및 로깅을 활성화하고 이벤트 발생 시 관리자에게 통지
I9	취약한 소프트웨어/펌웨어 (Insecure Software/ Firmware)	패치파일이 암호화되어 있지 않거나 민감정보를 포함, 업로드 전 검증과정을 거치지 않아 보안위협에 노출	사물인터넷 기기가 자동업데이트 기능을 제공하는지 확인, 패치파일에 대한 철저한 검증 등
I10	취약한 물리 보안 (Poor Physical Security)	USB 포트를 통한 기기 접근, 스토리지의 탈취 등	USB 포트 등 외부접근을 차단하고, 스토리지 탈취 시 데이터 리셋 등의 보호조치 마련

출처: 금융보안원(2016), 금융권 사물인터넷(IoT) 동향 및 향후 전망

다만 앞서 언급한 바와 같이 사물인터넷 환경에는 매우 많은 사물이 연결되어 있어 이를 관리하기 위해서는 사람에 의존하지 않을 수 있도록 자동화된 보안관리 기능을 갖출 필요가 있다. 특히 아래 표에 명시된 보안 기능 등을 고려하여 경량화·분산화 된 보안 솔루션을 갖출 필요가 있다. 특히 보안 수준은 해당 사물인터넷 환경을 통해 운영되는 서비스 및 데이터의 중요성에 따라서 다르게 설계할 필요가 있다. 이는 모든 환경에 동일한 보안 수준을 적용할 경우 필요한 알고리즘 및 연산량이 과도하게 증가할 수 있기 때문이다.

〈표 III-11〉 사물인터넷을 위한 보안 기능

사물인터넷 보안 기능	보안 기능에 대한 설명
사물인터넷 장치들 위한 보안 부팅 지원	<ul style="list-style-type: none"> <li>· 사물인터넷상에서 동작하는 각 장치들이 안전한 보안 연산 환경을 보장하기 위해서는 처음 스위치가 켜 졌을 때 펌웨어에 대한 인증값을 검증하여 무결성을 확인할 수 있는 보안부팅(secure booting) 기술이 필요</li> <li>· 이를 위해서는 운영체제 외적으로 별도의 장치에 의해서 전자서명과 같은 암호학적 연산이 동작될 수 있어야 함</li> </ul>
경량 암호 및 분산된 자발적 보안 설정 지원	<ul style="list-style-type: none"> <li>· 프라이버시 보호 및 암호화 방식은 단순하고 작은 장치에서도 적용 가능한 경량 암호화(lightweight encryption) 솔루션이 필요함</li> <li>· 최소 260억 개 이상의 사물들이 네트워크에 연결되기 때문에 보안 관리자에 의해서 모든 사물들에 대한 보안 파라미터를 적절하게 관리하는 것은 불가능</li> <li>· 따라서 관리자가 없이도 자발적으로 인증 및 보안을 위한 설정이 이루어지도록 해야 함</li> </ul>
사물들 간의 가상 사설망 설정 및 관리 지원	<ul style="list-style-type: none"> <li>· 공공 네트워크를 통해서 중요한 데이터를 전송하는 경우에는 사물들 간에 가상 사설망(virtual private network)을 설정하고 해제하는 것이 가능해야 함</li> <li>· 또한 각 장치들에 할당된 제한된 대역폭과 임베디드 장치의 간헐적 네트워크 연결 특징을 유지하면서 동시에 소프트웨어 업데이트와 보안 패치가 전달되는 매커니즘 또한 구성되어야 함</li> </ul>
빅데이터 분석에 대한 프라이버시 보호 기능	<ul style="list-style-type: none"> <li>· 사물인터넷상으로 많은 센서로부터 수집된 정보는 빅데이터 분석이 적용되는데 이때 프라이버시 보호기능이 제공되어야 하며, 사물인터넷 데이터에 대해서도 적절한 프라이버시 보호 기능 및 익명화 기술이 적용될 수 있음</li> </ul>
심층 패킷 정보감시 기능 지원	<ul style="list-style-type: none"> <li>· 심층 패킷 정보감시(DPI; deep pecket inspection)가 가능한 방화벽과 침입방지 시스템이 구성되어야 함</li> <li>· 필요에 따라 특정 장치를 목적으로 하는 트래픽에 대한 DPI 솔루션이 적용 되어야 함</li> </ul>

출처: 사물인터넷 시대의 사이버 물리 시스템 보안 기술 동향(연도)

추가적으로 고려할 사항은 사물인터넷 환경의 상당수를 차지하는 기기, 특히 센서의 성능이 그리 높지 않고 전력을 많이 사용하지 않는다는 점이다. 이에 다양한 경량형 암호화 및 메시지 프로토콜이 개발되고 있지만 전통적으로 고성능 기기에서 사용하던 보안기술에 비하여 알려지지 않은 약점이 있을 수도 있다. 예를 들어 많은 금융회사들이 도입하고 있는 비콘의 경우 원본 정보를 복제해 정보를 유출하는 클로닝(Cloning)이나 네트워크 통신 관련 정보를 속여 통신 흐름을 왜곡시키는 스푸핑(Spoofing) 등 일부 보안상 취약점이 존재하였었다. 따라서 이들 기술에 대한 충분한 검증 및 보안대책을 마련한 후에 도입을 검토하여야 한다.

또한 이와 관련 기존의 컴퓨터 등에 적용되던 보안솔루션에 비하여 센서에 적용할 수 있는 것은 전반적으로 단순하고 낮은 수준이 될 가능성이 높다. 이는 센서 자체가 소비할 수 있는 전력에 한계가 있고 자체 성능이 높지 않기 때문이다. 또한 센서와 같은 사물인터넷 기기에 적용된 보안솔루션은 이후 업데이트 상 한계가 있다. 이로 인해 보안솔루션에만 의존할 경우 회사 전체의 보안 수준이 낮아질 수 있다. 이에 사물인터넷을 도입하고자 하는 기업에서는 보안위험으로부터 스스로를 보호하기 위하여 물리적 보안방안 등을 추가적으로 강구할 필요가 있다.

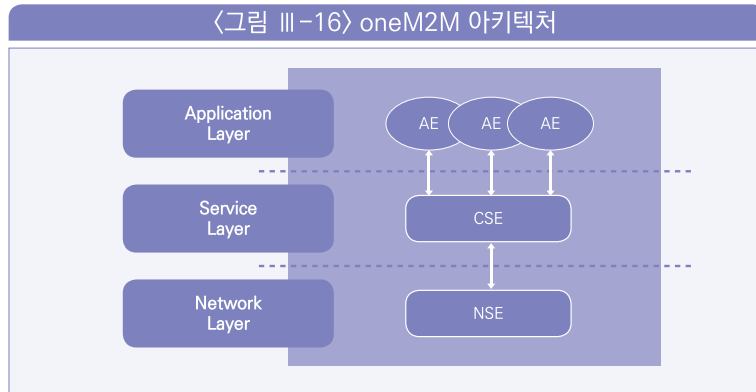
## 4 IoT 표준화 동향

현재 개발된 많은 IoT 서비스들은 특정 장치들에게만 제공되고 있으며, 동일한 서비스라도, 같은 제조사에서 만들어진 디바이스들 간에만 네트워크 통신이 원활하다. 따라서, 서로 다른 제조사의 사물들 간의 국제 표준 통신이 반드시 필요하다. 이에 컨소시엄 중심의 사물인터넷 플랫폼은 다국적의 다양한 업체들이 모여 업계 표준을 지향하여 공통적인 기술을 개발하고 있다. 현재 대표적으로 국제표준으로는 OneM2M(One Machine to Machine)와 OCF(Open Connectivity Foundation)이 있다.

### 4-1 oneM2M(글로벌 표준 플랫폼)

oneM2M은 사물통신, IoT 기술을 위한 요구사항, 아키텍처, API 사양, 보안 솔루션, 상호 운용성을 제공하는 글로벌 단체이며, 사양은 스마트 시티, 스마트 그리드, 커넥티드 카, 홈 오토메이션, 치안, 건강과 같은 다양한 애플리케이션과 서비스를 지원하는 프레임워크를 제공한다. oneM2M은 2012년 7월, TTA(한국), ATIS(미국), TIA(미국), ETSI(유럽), CCSA(중국), TDSI(인도), 전파산업회(ARIB, 일본), TTC(일본) 등 8개 주요 기관이 설립하였다.

현재 oneM2M은 200개 이상의 참여 파트너들과 회원들이 있으며, 여기에는 알카텔-루슨트, AT&T, BT 그룹, 어도비, 에릭슨, 도이체 텔레콤, IBM, 시스코 시스템즈, 시에라 와이어리스, 인터디지털, 인텔, LG유플러스, 텔레포니카 등 글로벌 유수의 기업이 포함된다. 특히 대한민국은 oneM2M 표준에 기반한 솔루션을 위한 주도적인 시장들 가운데 하나이다. 국내 이동통신 3사는 이미 oneM2M 플랫폼을 구축하였다. SK텔레콤은 ThingPlug 플랫폼에 oneM2M을 탑재하여 B2B를 중심으로 스마트 홈 등의 서비스에 활용하고 있으며, LG U+는 B2C 제품인 키즈와치를 출시하여 oneM2M 플랫폼을 활용하고 있고, KT는 IoTMakers 플랫폼에 oneM2M을 탑재하였다. 국내 기존 사물인터넷 관련 실증 사업으로 부산시, 대구시가 운영되고 있으며 추가적으로 고양시에 실증 단지가 추가될 예정으로 모두 oneM2M 기반의 사업으로 실증이 이루어지고 있다.



출처: 금융보안원, 금융권 사물인터넷(IoT) 동향 및 향후 전망(2016)

oneM2M 아키텍처는 애플리케이션(Application), 서비스(Service), 네트워크(Network) 레이어의 3개 계층 구조로 이루어져 있으며, 각 레이어는 AE (Application Entity), CSE(Common Service Entity), NSE(Network Service Entity)로 구성된다.

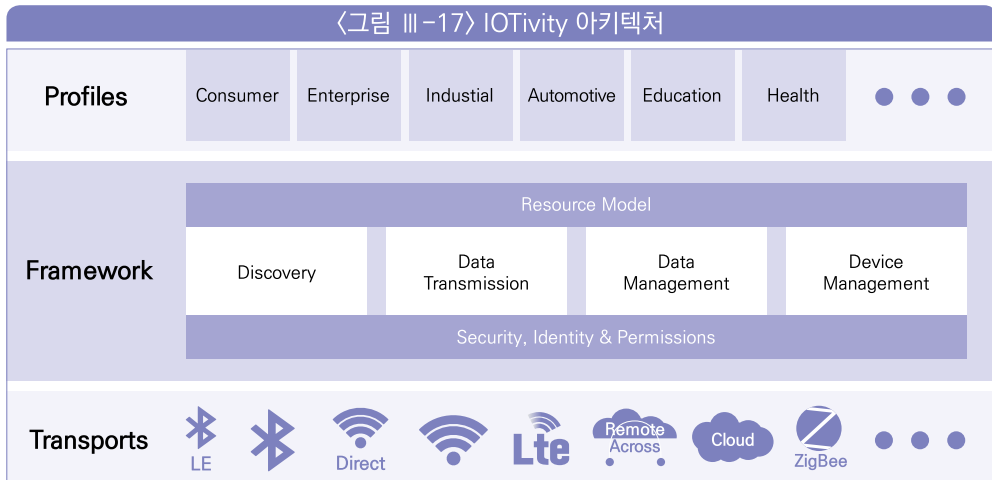
- **AE:** 다양한 서비스를 제공하는 애플리케이션을 의미
- **CSE:** 사물인터넷의 다양한 응용프로그램(AE)에게 공통적으로 제공해야 할 서비스 기능들(CSF, Common Service Function)을 포함한 계층으로, 장치 등록, 정보 탐색, 보안, 그룹 관리, 정보 구독·통지, 장치 관리, 데이터 저장·관리·분석, 메시지 관리, 서비스 과금 등 12가지 공통 기능을 제공
- **NSE:** CSE 간 통신이 가능하도록 통신 기능을 제공

#### 4-2 OIC의 IOTivity (산업단체 표준 플랫폼)

OCF(Open Interconnect Consortium)의 시작은 2014년 9월 삼성전자, 인텔, 시스코, 미디어텍 4개사가 창립 보드멤버로 설립한 OIC(Open Interconnect Consortium)이다. 여기에 UPnP포럼과 AllSeen Alliance 3개 멤버, 마이크로소프트, 퀄컴, 일렉트로룩스가

참여하였다. OCF는 오픈소스와 표준 규격을 동시에 개발하는 운영방법을 채택하여, 이를 통해 OCF의 표준을 빠르게 외부에 확산시키는 전략이다. 2015년 9월 표준규격 1.0을 발표했으며, 이후 릴리즈 버전 1.1.0을 공개한 상태이다. OCF의 오픈소스는 IoTivity이며 1.1.0 버전이 개발되어 공개되었다. 또한 기존 사물 간 연결에 더하여 추가하여 클라우드와의 연결을 지원하는 초안을 발표한 상태이다.

OCF에서 지원하는 IoTivity 플랫폼은 RESTful 아키텍처의 서버-클라이언트 방식으로 동작한다. 즉 플랫폼 내에서 연결된 모든 사물을 리소스로 구현하고, 사용자는 CRUDN(Create, Read, Update, Delete and Notification) 작동 명령어를 전달하여 플랫폼 내 리소스에 대한 조작 및 실행 등의 요청을 할 수 있다. 또한 플랫폼 내 새로운 사물이 발견되면 해당 사물에 대응한 IoTivity 리소스를 자동으로 생성하여 저장하게 된다.



출처: 금융보안원(2016), 금융권 사물인터넷(IoT) 동향 및 향후 전망

이러한 리소스 방식을 구현하기 위하여 IoTivity는 응용프로그램을 담당하는 프로파일(Profiles), 블루투스, 와이파이 등 사물인터넷의 다양한 통신 기능을 제공하는 전송(Transports), IoTivity 리소스 발견, 데이터 전송, 데이터 및 디바이스 관리 등 플랫폼의 주요 기능을 담당하는 프레임워크(Framework)의 3개 계층 구조로 이루어져 있다. 프레임워크의 주요 기능은 다음과 같다.

- **리소스 발견(Discovery):** 근접해있거나 원격에 있는 장치 및 자원을 발견
- **데이터 전송(Data Transmission):** 메시지 또는 스트리밍 기반으로 정보를 주고 받는 기능 제공
- **데이터 관리(Data Management):** 센서를 통해 수집한 정보 또는 다른 장치로부터 전달받은 정보를 저장, 분석, 활용
- **디바이스 관리(Device Management):** 장치의 구성, 권한설정 및 진단 등의 기능을 제공
- **보안, 인증 및 권한관리(Security, Identity & Permissions):** 프레임워크 전반에 걸쳐 보안, 인증 및 권한관리 등 보안기능 제공

IoTivity 서비스는 기본적인 사물간의 연동을 위한 근본적인 기능이라기 보다는IoTivity 기반 상용화를 보다 쉽게 지원하기 위한 추가적인 서비스 프레임워크이다. IoTivity 서비스는 크게 기본 서비스 블록과 추가 서비스 블록으로 구분되며, 기본 서비스 블록은 프로토콜 플러그인 관리자, 소프트웨어 센서 관리자, 사물 관리자, 알림 관리자 기능을 지원한다.

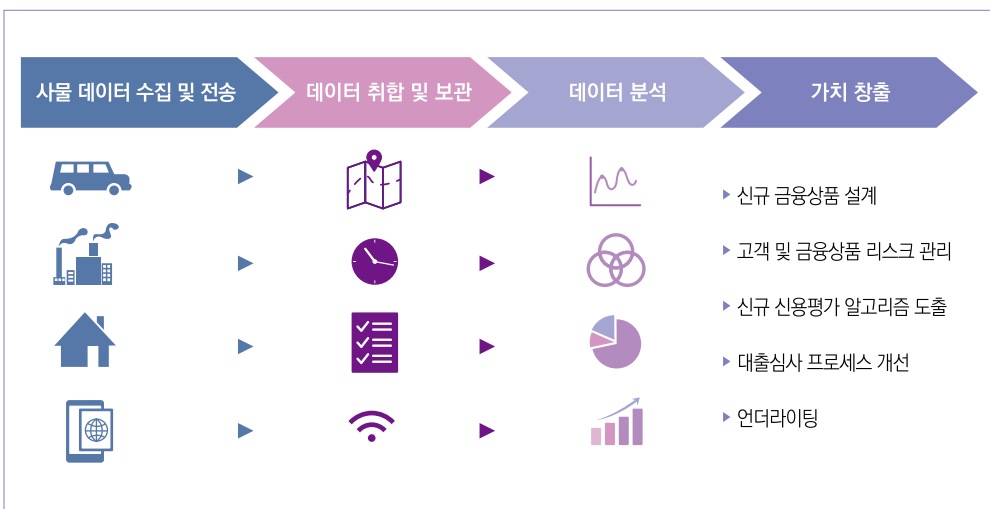
- **프로토콜 플러그인 관리자(PPM):** PPM은 IoTivity에서 지원하지 못하는 추가적인 프로토콜이 필요한 경우 정형화된 플러그인 기반 프로토콜 확장하여 사용할 수 있는 기능을 지원한다.
- **소프트웨어 센서 관리자(SSM):** SSM은 물리적, 논리적 센서 정보 등을 조합하여 새로운 가상의 센서 정보를 정의할 수 있는 소프트웨어 서비스 기능을 지원한다.
- **사물 관리자(TM):** TM은 사물을 원하는 형태로 그룹화하여 효과적으로 관리할 수 있는 기능을 지원한다. 이를 위해 기본적으로 사물에 대한 그룹 생성, 조회, 수정, 삭제 기능을 지원하고 그룹 상태나 디바이스 변경을 확인할 수 있는 기능을 제공한다.
- **알림 관리자(NM):** 특정 사물의 상태변화가 발생한 경우 이를 다른 디바이스로 알려주는 기능을 쉽게 개발할 수 있도록 지원하는 NM 기능을 지원한다.



1 핀테크 산업에서 IoT의 사용 현황

4차 산업혁명과 더불어 금융산업에도 다양한 혁신과 변화가 이뤄지고 있다. 소비자들은 온라인 상에서 통합적인 금융서비스를 원하고 있으며, 금융기업들은 이러한 소비자들의 요구사항에 따른 맞춤형 서비스를 제공하기 위하여 새로운 기술들을 도입하고 서비스에 적용하고 있다. 이러한 새로운 기술들에는 IoT, 데이터마이닝, 인공지능, 블록체인, 생체인식기술 등이 있다. 이 중 IoT는 광범위한 잠재력을 가진 기술로서 기존의 금융서비스나 신규 서비스에 통합하여 금융 위험을 줄이고 고객 가치를 향상시킬 수 있는 기술이다. IoT는 디바이스로부터 실시간으로 수집되는 다량의 데이터 처리가 가능하므로 거래의 편의성을 대폭 강화시킨 새로운 서비스 개발이 가능하다.

〈그림 III-18〉 핀테크에서의 사물인터넷을 활용한 가치 창출



이에 향후 IoT는 은행, 보험, 지급결제 등 다양한 금융 분야에서 폭넓게 활용 가능할 것으로 전망된다. 은행업은 IoT를 활용해 대출심사 프로세스를 개선할 수 있다. IoT를 통해 개인에 대한 생체 및 위치 센서 데이터를 분석하면 보다 안정적인 신용평가와 대출심사가 가능하기 때문이다. 카드사나 증권사의 경우도 고객들에게 맞춤형 상품정보를 발송하기를 원하는 경우 네트워크 기기에서 실시간으로 생성된 데이터를 취합해 고객에게 맞는 메시지로 가공해서 보내주는 메시지 사이트를 활용할 수 있다.

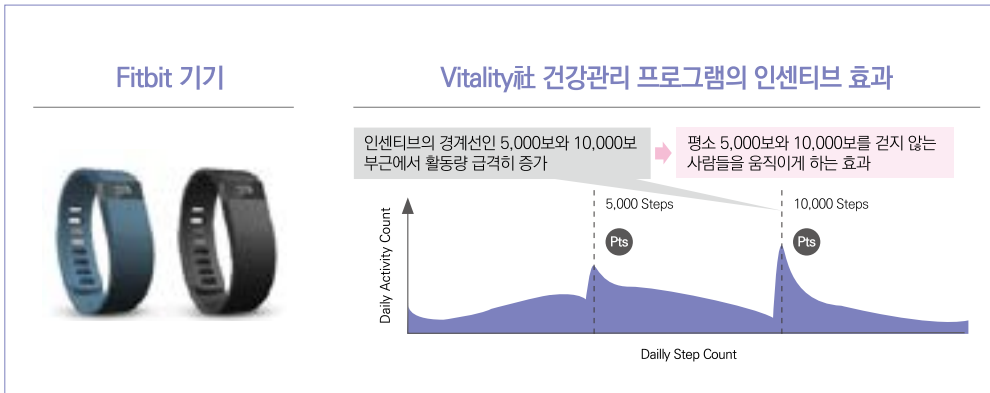
보험업에서도 IoT를 상품설계, 언더라이팅, 리스크관리와 보험료 책정 등에 활용할 수 있다. 실시간 IoT 데이터를 수집하여 개인의 행동을 파악하면 리스크평가 항목과 보장 항목을 상세하게 분류하여 고객맞춤형 보험 상품 개발이 가능하다. 뿐만 아니라 온라인과 오프라인의 결합을 의미하는 O2O가 등장하면서, IoT 기술을 활용한 결제시스템 도입 등 새로운 금융서비스가 확산되고 있는 추세다.

## 1-1 보험 산업과 IoT

보험 회사는 IoT를 채택하여 보상 정책, 보험료 및 고객을 포괄적이고 쉽게 액세스할 수 있는 데이터베이스에 통합함으로써 적용 범위 비용을 낮출 수 있다. 예를 들어, 자동차 보험에서 IoT는 추적 장치를 통해 보험에 가입한 실제 마일리지 및 운전 행동을 계산하는 데 사용된다. 주택 보험에서 스마트 임베디드 장치는 화재, 누수 또는 높은 수준의 일산화탄소를 감지할 수 있으므로 보험 회사의 위험이 낮아짐에 따라 보험료가 낮아질 수 있다.

건강관리에서 애플 또는 삼성과 같은 대기업은 정교한 모니터링 장치를 통해 개인의 건강을 추적하고 있으며 바이탈리티(Vitality)는 보험 계약자가 수행한 활동 수준에 따라 보상을 제공한다. 보험 부문에 IoT를 통합하고 핀테크/인슈어테크 신생 기업의 수가 증가함에 따라 농업 보험, 자동차 보험 및 건강 보험을 포함한 새로운 분야에서 이 부문의 성장이 더욱 가속화될 것으로 예상된다.

〈그림 III-19〉 Fitbit 웨어러블 기기와 Vitality社 건강관리 프로그램 효과



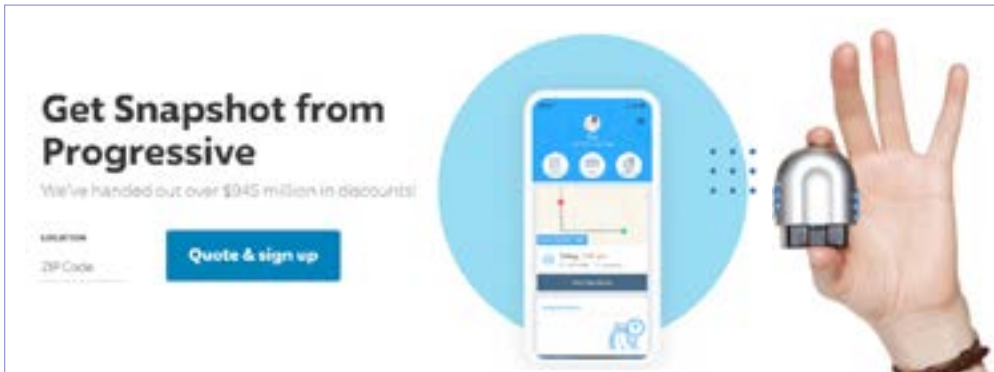
출처: Fitbit, Vitality(2014), "Wearables as Work"

이 장치는 운전자의 운행 데이터를 수집하고 운전 성향 등의 데이터를 AT&T 무선 통신을 통해 프로그래시브의 서버로 전송한다. 이 보험사는 운전자의 데이터를 진단하고 보험료를 깎아줄 수 있다. 보험료는 운전자의 성향에 따라 최대 30%까지 할인된다.

미국의 대형 자동차 보험사인 프로그래시브는 차량 운행기록 장치인 '스냅샷(Snapshot)' 프로그램을 제공하고 있다. 스냅샷은 차 내부에 사물인터넷 센서 및 통신 장치를 설치해 운전자의 운전 습관을 파악한 후 모범운전자에게는 보험금 할인 혜택을 제공하는 프로그램이다. 보험사는 이 스냅샷 장치를 통해 가입자 자동차의 주행거리는 물론 주행속도, 급제동 및 급가속 여부 등 운전자의 운전 성향에 대한 정보를 수집할 수 있다. 이를 위해 운전자는 먼저 프로그래시브에서 제공한 스냅샷 장치를 차에 30일 동안 부착한다.

프로그래시브는 데이터를 진단하는 과정에서 빅데이터 기술을 활용해 정보를 분석하고 각각의 가입자의 운전 위험성에 따라 차별적인 보험료를 책정하고 있다. 이를 통해 보험사는 수익성을 달성하는 동시에 보험가입자의 비용을 절감할 수 있어 고객만족도 또한 높일 수 있다.

〈그림 III-20〉 프로그레시브의 스냅샷

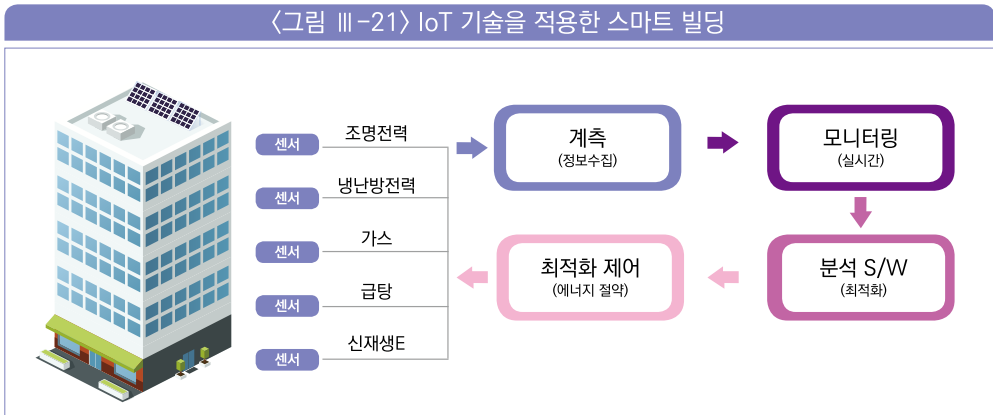


출처: progressive.com

미국 여행자 보험사인 더 트레블러스 인데미티 컴퍼니(The Travelers Indemnity Company)는 보험상품과 IoT 기술을 결합한 신상품을 선보였다. 자사의 보험을 가입하는 고객에게 경보용 센서를 제공하여 고객이 장기 출타 혹은 여행 중이더라도 집안에서 벌어지는 화재 및 도난, 기타 이상 상황을 쉽게 인지할 수 있도록 경고 알림을 지원한다. 이를 통해 트레블러스는 자사의 보험상품에 스마트홈 구축이라는 패키지 형태의 상품을 판매할 뿐만 아니라 고객의 데이터를 수집하여 더 나은 보험상품을 개발하는 데 활용할 수 있게 된다. 또한 고객 입장에서는 단순한 보장상품으로서의 보험 가입 뿐만 아니라 기술적 지원을 통하여 사고를 사전에 예방할 수 있게 되어 장시간 안심하고 출타가 가능하게 된다.

## 1-2 부동산과 IoT

핀테크 스타트업은 사물 인터넷을 사용하여 상업용 부동산에 새로운 기회를 창출하고 있다. IoT 기술을 통해 스타트업은 수리, 유지 보수 및 에너지 사용에 대한 데이터를 수집하여 자산 가치를 보다 잘 평가할 수 있다. 또한 센서를 통해 에너지 비용을 조절할 수 있으며 특히 산업 단지 내 건물 유지 보수를 최대 30%까지 줄일 수 있다. 이 데이터가 공개되면 거래 시장이 생겨 입찰 시 투자자에게 더 높은 투명성과 정확성을 제공할 수 있다. 이는 부동산에 IoT가 관여하면 실시간 입찰 시장에 혁명을 일으킬 수 있음을 의미한다.



출처: 에너지관리공단(2020)

### 1-3 O2O와 IoT

알리바바와 아마존과 같은 전자 상거래 거대 기업은 기존 소매 유통업을 흔들기 위해 열심히 노력했다. 그러나 얼마 지나지 않아 온라인에서만 운영하는 것이 장기적인 지배력을 발휘할 수 없다는 것을 깨달았다. 아마존은 2000년 아마존 프레스리 스토어를 오픈하기 시작했으며 알리바바는 중국의 할인 슈퍼마켓에 3억 달러 이상을 투자하고 있다.



출처: 아마존(2021)

O2O(Online-to-Offline) 상거래는 온라인 상점에서 오프라인 비즈니스로 고객을 전환하기 위해 일반적으로 사용되는 비즈니스 전략이다. O2O는 인터넷 광고 및 이메일 마케팅 캠페인을 통해 온라인에서 잠재 고객을 식별하고 오프라인 비즈니스로 전환한다. 따라서 실제 제품을 보고 싶어하는 고객은 온라인 상점을 방문하여 원하는 제품을 선택하고, 실제 구매는 오프라인 매장에서 이루어진다. 온라인과 오프라인 모두에 존재하는 회사는 교차 판매 수단으로 O2O 비즈니스 전략을 사용한다.

핀테크 산업에서 O2O는 점점 인기를 얻고 있다. 이미 시장에서도 확인된 바와 같이, IoT 관련 통계는 많은 인구, 신속한 도시화 및 높은 스마트폰 보급률 증가로 인해 동남아시아에 가장 유리할 것으로 보인다. 인프라 측면에서 중국, 인도네시아, 싱가포르 및 한국은 핀테크 허브가 되기에 유리한 환경이다. 중국은 인구가 많고 기술 사용에 대한 개방성으로 인해 O2O의 선두 주자이다. Click-and-Collect 서비스(온라인으로 주문하고 오프라인 상점에서 찾기로 한 주문)에 주로 중점을 둔 영국 및 미국 시장과 달리 중국의 O2O 마켓에는 주문형 서비스, 일일 거래 사이트 및 Click-and-Collect 서비스가 포함된다. 중국 소비자는 쿼이디(Kuaidi)의 택시 앱을 사용하고 택시비 60% 할인을 받거나 58 다오지아(58 Daojia)의 홈 서비스를 사용하여 리베이트를 받을 수 있다. 중국의 O2O 시장은 여행, 운송, 세탁 서비스, 음식 배달 등 온라인 식료품 쇼핑을 넘어서었다.

#### 1-4 간편 지급결제 서비스와 IoT

마스터카드사는 단순한 플라스틱 카드회사를 넘어 지급결제 시장의 선도적인 플랫폼사로 진화하고자 다양한 시도를 하였다. 이를 위해 삼성전자 및 전자상거래 업체인 프레시다이렉트(FreshDirect), 쇼라이트(ShopRite)와 협력하여 새로이 그로서리즈(Groceries) 앱을 개발하였다. 이 앱을 통하여 소비자는 별도의 PC나 모바일기기에 접속할 필요없이 냉장고에 남은 식료품 양을 보고 삼성에서 개발한 냉장고의 터치 스크린을 통해 필요한 식료품을 즉시 주문하고 결제 할 수 있게 되었다. 고객이 주문한 식료품은 IoT 기술을 통해 즉시 주요 식료품 업체로 주문이 진행된다.

〈그림 III-23〉 삼성 Family Hub 냉장고와 터치스크린



출처: 삼성

### 1-5 대출 서비스와 IoT

일본의 핀테크 기업 GMS(Global Mobility Service)는 필리핀에서 IoT를 활용한 신용서비스를 제공하고 있다. 필리핀에서는 다수의 3륜차 트라이시클이 운행되고 있다. 하지만 트라이시클을 구매하려고 해도 개인신용 문제로 대출이 되지 않아 구매를 할 수 없는 가난한 사람이 많다. 이러한 빈곤층들은 트라이시클을 고액으로 임차할 수 밖에 없어 빈곤한 상황이 지속되는 악순환이 발생한다. GMS는 이러한 대출사이클의 악순환을 해결하고자 신규 대출을 차주들에게 제공하고 있다.

〈그림 III-24〉 GMS의 IoT 장비 MCCS와 이 장비가 장착된 오토바이



출처: GMS 웹사이트, Kawasaki

차주는 우선 GMS를 이용해 대출받아 모빌리티(자동차, 트라이시클 등)를 구매할 수 있다. 이렇게 구매한 모빌리티에는 GMS가 개발한 IoT 디바이스 'MCCS(Mobility-Cloud Connecting System)'를 부착해야 한다. 이 기기는 엔진 시동 등 원격제어관리 기능이 있어 해당 모빌리티의 주행 관련 데이터를 수집한다. GMS는 이렇게 수집한 데이터를 차주의 일반 정보와 결합하고 분석하여 지불능력 등을 수치화해 새로운 신용정보로 평가한다. 만약 이 분석결과가 좋지 않은 경우에는 지불능력이 없다고 판단해 구매한 모빌리티의 엔진 시동을 걸 수 없게 원격조작할 수 있는 방식이다.

이를 통해 차주가 대출 상황을 지키도록 유도하여, 대출 미상환에 대한 리스크를 낮추고 있다. 한편, 체납이 발생할 경우 이자 등을 다시 상환하면 바로 엔진이 재가동 할 수 있으며, 도로에서는 엔진이 정지하지 않도록 안전기능 갖추고 있다. 이러한 방식을 통해 이전에는 대출 대상 외로 판단됐던 고객이 대출이 가능해졌으며, 채무불이행률은 1% 이하로 매우 낮은 수치를 기록하고 있다. GMS는 현재 필리핀을 시작으로 캄보디아, 인도네시아 그리고 일본에서 사업을 전개하고 있다.

〈그림 III-25〉 GMS의 비즈니스 모델



출처: GMS 웹사이트



## 2 금융에서의 IoT 도입 방법

금융산업은 이미 IoT 기술을 사물에 대한 데이터에 직접 연관된 비즈니스의 요소를 측정, 분석하는 데 사용하고 있다. 예를 들어 자동차보험 가입자의 운전 습관, 건강 보험 가입자의 건강 상태 등이 그 예이다. 따라서, 금융산업에서의 가까운 미래의 IoT 잠재력은 이러한 기존의 “실체적인” 적용 방안이 얼마나 확산되느냐에 달려 있을 것이다.

금융기관은 사물 센서에서 얻는 데이터를 활용하여 기존의 기회를 보다 향상시킬 수 있을 것이다. 예를 들어, 자동차 내 센서 장착 증가는 보험사가 운행량 기반 보험의 도입을 추진하는 데 쓸 수 있는 데이터를 제공할 것이다. 소비자 부문에서는 스마트홈 통제를 위한 상당한 규모의 센서 설치가 이뤄질 것으로 예상되는데, 이는 금융산업에도 도움이 될 수 있다. 부동산 담보대출 심사에서 집의 상태를 파악해 담보가치를 평가하는 데 이용할 수 있고, 화재보험의 리스크 관리와 정확한 보험료 측정에도 이용이 가능하다. 상업 부문에서는 소비재 제조업체와 농·목축업 활동 및 현장 모니터링을 위한 센서를 금융기관들이 사용할 수 있다. 증권사나 기업금융 은행은 이들 센서가 생성하는 데이터를 투자 혹은 대출 활동에 이용할 수 있다. 운송 중인 상품에 부착된 센서는 은행의 현금 관리 및 거래 서비스 사업에 기회를 제공할 수 있다. 상품 판매자와 구매자 간의 지급결제 자금 흐름을 보다 원활하게 조율해줄 수 있기 때문이다.

〈그림 III-26〉 IoT 기술을 적용한 커넥티드 차량 구상도



출처: utoimage (2021)

외부 제공 서비스와는 별도로, 금융기관은 IoT를 이용해 내부 업무 개선에 사용할 수 있다. 단순히 직원들의 위치와 움직임에 대한 정보를 파악하는 것만으로도 직원들의 업무, 상호작용, 아이디어 공유에 대한 풍부한 인사이트를 얻을 수 있다.

## 1 IoT에 의한 핀테크 미래 혁신

엑센츄어(Accenture)는 가까운 미래에 사물인터넷이 금융분야로 확산되어 금융회사를 이용하는 고객들의 생활에 자연스럽게 동화될 것으로 예상하고 있으며, 고객들 또한 언제, 어디서, 어떻게 그들이 원하는 서비스를 받을 수 있을지 잘 알고 있을 것으로 보고 있다. 이러한 환경에서 금융회사는 사물인터넷으로부터 수집한 정보들을 활용하여 자신들의 고객을 보다 더 잘 이해할 수 있도록 끊임없이 노력해야 하며, 결국엔 금융회사의 이러한 노력이 고객의 삶을 변화시킬 것으로 보고 있다. 엑센츄어는 이러한 금융회사에 특화된 사물인터넷 환경을 Bank of Things라는 용어로 정의를 하고 있다.

Bank of Things 환경에서의 금융회사는 다양한 사물로부터 수집한 정보를 이용하여 고객의 금융·비금융 관련 니즈에 만족할 수 있는 개인 맞춤형 조언을 제공(advice provider)하게 될 것이며, 타 업종과 파트너십을 맺어 특정 고객에게 가치 있는 정보를 제공(value aggregator)하게 될 것이다. 또한 고객을 다른 서비스 제공자와 연결(access facilitator)해줌으로써 고객의 삶을 변화시킬 것이다.

이와 같이 Bank of Things 환경에서 고객의 삶을 변화시키는 금융회사의 다양한 역할에 대해 엑센츄어는 세 가지(개인뱅킹, 기업뱅킹, 산업뱅킹) 미래 가능 시나리오를 제시하고 있다.

**(개인뱅킹 시나리오)** Yumi Sato는 퇴근길에 차가 고장났다는 경고등을 확인했다. 그녀는 자동차 수리비가 얼마나 나올지 걱정이 되어 집에 도착하자마자 스마트폰의 뱅킹 앱을 실행하여 계좌잔고가 얼마 남지 않은 것을 확인하였다. 뱅킹 앱은 집 안의 스마트 냉장고, 전기계측기 및 수조 등 스마트홈 기기들로부터 자료를 자동으로 다운로드받아 현재 소비한 전기세, 수도세 등을 실시간으로 계산하여 Yumi의 자산을 파악한 후 몇 가지 제안을 한다. 한 가지는 휴가를 위해 6개월간 모은 적금을 자동차 수리비에 사용하거나, 다른 한 가지는 신용카드 한도를 높이는 것이다. 마지막으로 뱅킹 앱은 그녀의 운전습관을 분석한 데이터를 토대로 자동차보험이 줄어든 것까지 계산하여 보여준다. Yumi는 적금을 수리비에 이용하기로 결정했으며, 뱅킹앱은 다시 여행자금을 위한 적금을 들 수 있는 가이드를 제시한다.

**(기업뱅킹 시나리오)** Hawkins Weill(H+W)은 세계적인 의료기기 제조업체이다. H+W에게 있어서 제품이 생산되어 소비자에게 전달되기까지의 공급망을 관리하는 것은 기업의 수익과 직결되는 중요한 문제이다. 따라서 공급망 단계별 데이터를 수집하여 관리해줌으로써 재고를 줄이고, 수익을 극대화할 수 있다.

H+W와 거래하는 은행 또한 H+W 제품의 공급망에 대한 데이터를 수집하여 대차대조표, 상품회전률 등을 보다 면밀히 분석함으로써 H+W에 대한 투자를 동적으로 결정할 수 있다. 몇 달 전 은행은 H+W에서 현금흐름 문제가 발생할 수 있다는 것을 예측하고 선제적으로 H+W에게 대처할 것을 제안하였다.

**(산업뱅킹 시나리오)** John Martin 가족은 100년 이상 밀을 재배해 왔다. 다양한 센서들이 그의 농장의 토양, 농작물, 축산물 및 농기계 상태들을 매일 업데이트 해준다. John은 주거래 은행에 센서로부터 얻은 데이터들을 전송하면 은행은 농장 부동산 가치, 농작물의 잠재적 수익률 등을 계산하여 대출가능 금액 등을 제시해준다. 지난 몇 년간의 은행의 분석은 John이 농장을 경영하는 데 큰 도움을 주었다.

또한, 딜로이트는 금융분야 업권별 사물인터넷의 다양한 활용 시나리오를 제시했다.

## 1-1 은행권(Banking)

사물인터넷은 향후 은행의 보증 절차를 개선하고, 새로운 시장에 진출하는 등 은행에게 다양한 기회를 제공할 수 있을 것으로 전망하고 있다. 바이오 센서 및 위치 센서가 수집한 사람의 신체적, 행동적 정보 또는 기업의 제조, 운송 정보를 분석함으로써 개인 또는 기업의 신용평가가 가능할 것이다. 하지만 이러한 신규 기술을 도입하기에 앞서 무수히 많은 사물인터넷 정보들 중 어떤 정보가 신용평가에 적절한지를 분석하고, 사람의 신체적, 행동적 정보수집으로 인한 사생활 침해 관련 잠재적 위협을 어떻게 해결할 수 있을지를 우선적으로 고민해야 한다.

또한 은행은 물품 구입/대여 관련 금융상품 운용 시 물품에 부착된 센서로부터 물품의 상태를 모니터링함으로써 자산의 잔여가치를 정확히 분석할 수 있다. 모니터링을 통해 수집된 자산 상태에 따라 구입/대여한 이용자에게 할인을 적용하거나 위약금을 부과할 수 있을 것이다.

## 1-2 자본시장(Capital Market)

자본시장에서 사물인터넷 센서로부터 수집된 정보를 분석한다면 트레이딩 및 투자 활동을 자동화하는 데 더욱 도움을 줄 것으로 예측하고 있다. 또한 인간의 개입을 배제하고 사물인터넷 기반의 실시간 데이터 흐름을 분석함으로써 기업은 시장거품 여부를 더 정확히 평가할 수 있을 것으로 전망했다. 물론 일부는 자동화된 인공지능이 투자자의 요구나 지정학적 사건의 변화를 확실히 감지하지 못하기 때문에 잘못된 결정을 내릴 수도 있다고 보고 있다.

## 1-3 보험업(Insurance)

보험업권은 이미 센서로부터 수집된 정보들을 이용하여 자동차 보험에 적용하는 등 사물인터넷을 널리 활용하고 있다. 하지만 얼마전 발생한 테슬라의 자동주행 사고사례와 같이 자동주행 자동차의 보급으로 운전자가 책임을 지는 자동차 사고보험에서 자동차 제조사가 책임을 지는 생산물배상책임보험으로 변화할 것이다. 이 경우 보험사들은 사물인터넷을 통해 자동차의 결함 관련 정보를 보다 정확하게 수집하는 것이 가능할 것이다. 하지만 자동차 사고율의 급감과 전통적인 보험담보 적용이 힘들어져 보험료 수익이 상당히 줄어들 것으로 전망된다.

또한 UBI(운전습관연계보험)로 인해 보험계약자들은 그들의 보험료를 낮추기 위해 다양한 요구를 하게 될 것이다. 예를 들어 현재는 개인상해보험에서 모든 종류의 리스크가 하나의 약관에 포함되어 있지만 사물인터넷을 통해 개인 행동에 대한 다양한 정보를 수집할 수 있다면 보험사는 개개인에 대한 맞춤형 보험 상품을 개발할 수 있게 될 것이다. 이러한 제품의 차별화는 고객 만족도를 더욱 증가시켜 줄 것으로 예상된다.

상업의 경우 화물 컨테이너나 운송 차량에 센서를 부착하여 도난, 파손 등의 정보를 수집함으로써 보험사들은 운송보험을 개선할 수 있을 것으로 기대된다. 다양한 정보를 이용하여 보험사들은 자산위험을 보다 정확하게 평가할 수 있으며 따라서 보험료 책정도 보다 명확해질 수 있다.

#### 1-4 투자 및 자산관리(Investment and wealth management)

금융기관은 사물인터넷으로부터 수집한 고객정보(행동, 선호도, 위치정보 등)를 이용하여 투자결정과 자산할당 등에 활용할 수 있다. 예를 들어 고객의 관심분야, 구매패턴 등을 더욱 면밀히 분석함으로써 고객의 자산관리 개선 및 투자제안에 고객정보는 많은 도움이 될 것이다. 또한 기존의 설문조사 등을 통해 분석하였던 고객의 리스크 수용도를 사물인터넷 수집 정보를 프로그래밍화(POL-based)하여 분석함으로써 훨씬 정확한 결과 도출이 가능할 것이다.

또한 다양한 센서로부터 수집되는 실시간 정보를 이용하여 포트폴리오 관리를 자동화함으로써 투자관리 운용사, 펀드, 가격산정 전략을 차별화할 수 있을 것으로 보인다. 이러한 자동화된 기술을 이용하여 방대한 양의 데이터를 종합하여 트레이딩 기술에 융합함으로써 현재 인간이 수행하는 것보다 훨씬 빠르게 대응할 수 있다.

#### 1-5 상업용 부동산(Commercial real estate)

상업용 부동산에 실시간입찰(RTB) 시장이 출현함에 따라 전문가들은 다양한 시나리오를 구상할 수 있게 되었다. 이미 부동산을 검색하고 임대해주는 과정을 투명하게 제공해주는 서비스 관련 스타트업들이 출현하고 있으며, 빌딩의 에너지 및 보안 관련 데이터를 수집하는 센서, 엘리베이터 등 건물 주변 환경과 사람 간 상호작용에 관련된 데이터를 수집하는 센서 등을 활용하여 부동산의 가치를 더욱 정확하게 평가할 수 있을 것이다. 또한 투자자들에게 보다 투명한 가격을 제공함으로써 부동산 임대나 매매 시의 불화를 줄여줄 수 있을 것이다.

또한 거주자의 행동을 분석하고 건축 자재를 모니터링하여 얻은 데이터는 상업 및 주거용 건물을 디자인하고 건축할 때도 활용이 가능하다. 예를 들어 건설장비에 부착된 센서로부터 얻은 데이터는 건설회사가 프로젝트 진행 시 안전사고를 예방하는 데 도움이 될 것이다.

#### 1-6 리스크 관리(Risk management)

사물인터넷은 금융회사의 다양한 잠재적인 리스크를 감소시키고 성과를 향상시키는 데

도움이 될 것이다. 예를 들어 금융회사 직원들의 스트레스 레벨, 이동패턴 등을 모니터링함으로써 잠재적인 내부 부정행위를 예측하여 행위 리스크를 보다 잘 관리할 수 있다.

또한 가상 또는 현실 환경에서 다중요소(multi-factor) 인증을 사용하여 인증 도용을 방지할 수 있다. 예를 들어 소매상인은 카드를 이용하여 결제 처리 시 카드와 연관관계를 사전에 설정한 물리적 장비(스마트폰 등)가 근처에 있는지 여부를 검증하여 해당 고객이 정당한 사용자인지를 확인할 수 있다.

또한 포트폴리오 매니저들은 직원들이 스트레스를 받는 상황에서 그들이 어떻게 행동하는지를 분석하여 삶의 질을 개선시킬 수 있다. 물론 직원들은 본인들이 감시당하는 것을 싫어하겠지만 해당 직원들이 고객정보 등 중요정보를 취급하는 업무를 담당하는 경우, 이러한 감시가 가까운 미래에는 채용의 의무사항으로 반영될 수도 있을 것이다.

## 2 IIoT(Industrial IoT)의 소개

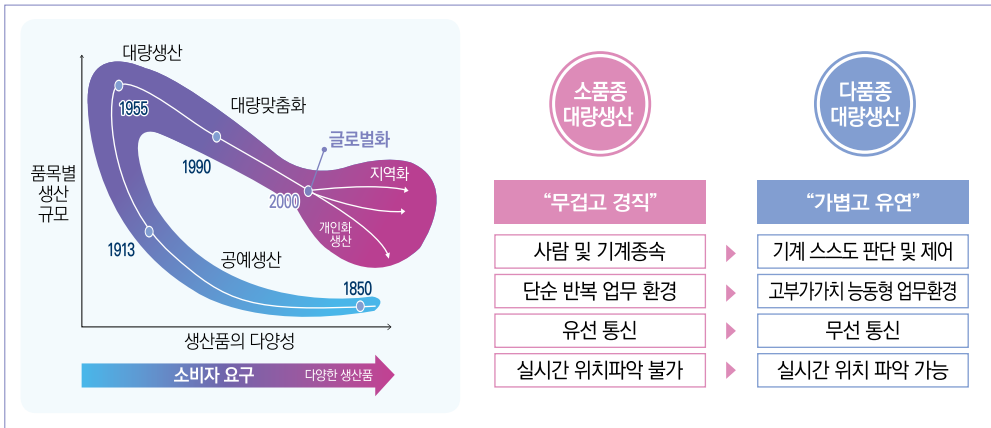
### 2-1 제조업의 변화

제조업은 향후 소비자의 다양한 선호에 맞춘 다품종 유연생산으로 패러다임이 변화될 전망이다. 미래의 제조업은 기존의 소품종 대량생산 체제에서 벗어나 다변화되는 제품 수요에 부합하는 모듈 공정 형태의 유연생산, 협동생산 등 다른 형태의 제조 형태가 등장할 것으로 예상된다.

또한, 제조업은 노동자, 자본 등 기존 요소 투입형 생산구조에서 향후 데이터 기반의 생산 구조로 전환하는 등 제조 방식의 근본적 변화가 예상된다. 제조업은 데이터 기반 디지털화가 진행됨에 따라 생산과정에서 기존의 사후 대응체계가 실시간 대응체제로 진화하고, 의사결정의 속도와 신뢰도가 향상됨에 따라 자율적이고 선제적 자원관리의 최적화가 이뤄질 것이다.

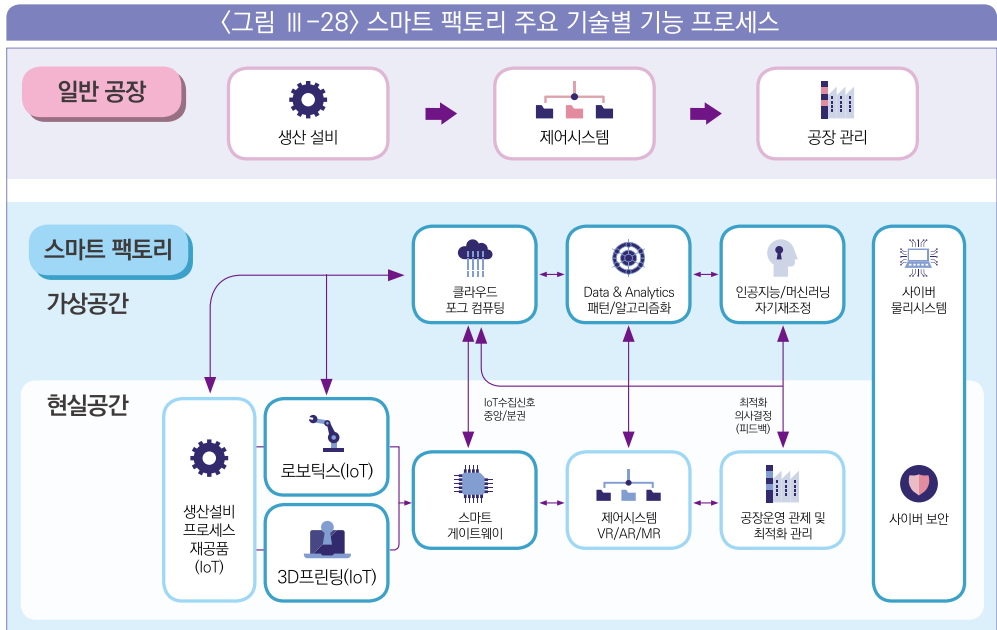


<그림 III-27> 제조업 벨류체인의 변화



맥킨지사는 사물인터넷 기술이 제조업에서 가장 큰 파급효과를 가져올 것으로 예측하였으며, 생산 공정에 도입함으로써 기업의 생산성 제고와 비용 절감에 크게 기여할 것으로 전망했다. 구체적으로 사물인터넷으로 인한 경제적 파급효과가 2025년까지 최대 11.1백만 달러 규모를 창출할 것으로 예측하였으며, 그중 제조 산업의 비중이 33.2%로 타 산업 대비 파급효과가 가장 클 것으로 전망했다.

또한, 제조 산업에 사물인터넷이 도입됨으로써 판매량 증가율이 연 5~15% 증가하고, 제조 비용 25%~35% 감소하며, 1인당 생산량이 40~60%가 향상하는 등 생산 효율성과 경제적 효과를 창출할 것으로 전망된다. 사물인터넷은 “스마트공장” 또는 “스마트제조”로도 언급되고 있는 제조 산업의 스마트화를 위해 현실 세계와 가상세계를 연결하는 접점의 역할 또한 하고 있다. 제조 스마트화는 사물인터넷을 통해 각종 생산설비에서 데이터를 생성·공유하며 인공지능은 공정 최적화 및 사후 예측을 수행하고, 궁극적으로 CPS(Cyber Physical System)를 통해 가상세계에서의 지능형 제조 서비스를 제공한다.



출처: 삼성KPMG 경제연구원(2018), 4차 산업혁명과 제조혁신:스마트 팩토리 도입과 제조업 패러다임 변화

## 2-2 IIoT 정의 및 범위

IIoT는 크게 소비자 IoT(CIoT; Consumer IoT)와 산업용 IoT(IIoT; Industrial IoT)로 구분되며 제조 산업에 사용되는 IoT는 고신뢰성을 요구하는 산업용 IoT에 해당한다. 각각의 IoT는 사물에 센서를 내장·부착하여 통신기술을 통해 실시간으로 데이터를 주고받는 프로세스는 동일하나, 산업용 IoT는 데이터의 응답 속도가 지연될 시 심각한 문제를 초래하기 때문에 높은 수준의 신뢰성을 요구한다.

〈표 III-12〉 IIoT 구분별 주요 특성

구분	소비자 IoT(Consumer IoT)	산업용 IoT(Industrial IoT)
개념	고객 서비스 중심으로 사물과 사물 간의 상호작용을 지원하며, 사용자와 밀접한 범위 내에서 주로 활용	산업분야에서 활용되는 사물인터넷으로, 산업 장비와 디바이스를 연결하는 데 중점
특징	소비자 중심, 일반적으로 시스템 오류 시, 위험한 상황이 일어나지 않는 분야에 적합	시스템 고장이나 오류, 중단/지연 상황에 심각한 위험 발생 가능한 분야에 적용되어 높은 수준의 신뢰성 요구
적용 범위	일상 생활, 주거, 오피스, 대형 마트 등 사용자와 밀접한 범위에서 편의성 향상을 목적으로 활용	제조, 운송, 석유, 화학, 플랜트 등 다양한 산업분야 전반에 기반기술로 활용

출처: KISTEP 기술동향브리프 (2020)

제조 공장의 스마트화 수준에 따라 사물인터넷 사용 범위가 다양하며, 현재 자율적인 공정 운영을 위한 고도화 단계의 기술 개발이 활발히 진행되고 있다. 제조업에서 IoT 사용 범위는 자재를 단순히 식별하여 점검하기 위해 사용되는 1단계 수준부터 공정 제어, 최적화까지 자율로 운영하기 위해 작업자, 설비, 자재, 운영 환경을 모두 연결하는 5단계 수준으로 구분 가능하다. IIoT에서의 연결 대상은 작업자(Man), 자재(Materials), 설비(Machinery), 운전조건(Methods), 환경(Environment) 등이 있다.

〈표 III-13〉 제조 산업에서의 사물인터넷의 사용 범위

	수준 정의	표준	IoT대상	특성	조건(구축수준)	주요 도구
고도화	Level 5	자율운영	작업자, 설비, 자재, 운전조건+환경	맞춤 및 자율 (Customized)	모니터링부터 제어, 최적화까지 자율로 진행	인공지능, AR/VR, CPS 등
중간 2	Level 4	최적화	작업자, 설비, 자재, 운전조건	최적화 (Optimized)	공정운영 시뮬레이션을 통해 사전 대응 가능	센서 제어기 최적화 도구
중간 1	Level 3	제어	작업자, 설비, 자재	분석 (Analysed)	수집된 정보를 분석하여 제어 가능	센서 + 분석도구
기초	Level 2	모니터링	작업자, 설비, 자재	측정 (Measured)	생산정보의 모니터링이 실시간 가능함	센서
	Level 1	점검	자재	식별 (Identified)	부분적 표준화 및 데이터 관리	바코드 RFID

출처: 스마트제조혁신추진단 홈페이지

## 1 NFC를 활용한 출입통제 시스템 만들기

NFC를 활용한 출입통제 시스템을 구현하기 위해서 관리자, 사용자 스마트폰 애플리케이션을 구현하고 Apache, PHP, MySQL을 이용해 서버와 데이터베이스를 구축하고 관리자 프로그램은 자바를 이용하여 구현할 수 있다. 개인 인증 분야에서도 기존 RFID 카드 대신 NFC로 대체되고 있는데 NFC를 활용한 출입통제 시스템은 스마트폰의 기능 중 하나인 NFC를 이용하여 사원증 소지의 불편함과 리더기 설치 시의 비용, 기존 리더기에 사용자 정보를 저장함으로써 악의적인 해킹에 노출되어 있는 문제를 해소하고자 출입통제 시스템을 설계 및 구현한다.

### 1-1 NFC

NFC는 기존 RFID의 확장된 개념으로 Sony와 NXP가 공동 개발한 개방형 RFID 기술의 일종으로 이동통신 단말기에 탑재하여 다양한 부가서비스가 가능하고 13.56MHz 주파수 대역을 사용하여 비접촉식으로 10cm 이내의 단거리에서 기기 간 데이터를 전송하는 근거리 무선통신기술이다. 기존의 비접촉식 스마트카드 기술(ISO/IEC 14443 Proximity-card 표준)을 기반으로 개발되었다.

〈그림 III-29〉 NFC 리더기



출처: acs.com

신용카드, 신분증 등을 대체할 수도 있으며, 노트북의 사용자 인증, 모바일 티켓, 쿠폰 등 다양한 분야에서 활용될 수 있는 성장 잠재력이 큰 기술이다. 정보의 읽기뿐 아니라 쓰기도 가능한 양방향 통신을 지원한다. 또한, 정보의 Encryption으로 타 RFID보다 안정성이 뛰어나다. 최근, RFID 기술의 발달로 PassiveType과 ActiveType을 포함해서 최대 900m 까지 사용 거리가 늘어났으며 리더와 태그 개념을 적용해 비즈니스 모델에 따라 활용 범위가 크게 확대될 것이라고 전망된다. 실생활과 연계된 복잡한 정보활동에 대한 해결책으로, 모든 타입의 사용자 장치에 대한 'Touch-and-Start' 조작으로 접촉과 실행에 대한 직관적 접촉 편의성을 제공하여 보안 처리가 내장된(Built-Insecurity) 지불/재정 등의 응용을 사용자들이 쉽게 사용할 수 있게 하였으며 스마트폰, AV(Audio Visual) 장비, 디지털 카메라, PDA, 세트 탑 박스 등과 컴퓨터를 직관적으로 접속하게 하여, 전화번호, 그림, 티켓, MP3파일, 북마크 등의 모든 콘텐츠를 용이하게 동작시킬 수 있다.

능동모드 혹은 수동모드에서 운영될 수 있어, 비접촉 스마트카드나 RF 태그와 같은 다양한 수동형 장치와 폭넓게 통신할 수 있다. 또한, 휴대폰, 블루투스 그리고 무선 802.11 장치에 대한 무선 네트워크의 자동 설정과 무선통신 서비스를 제공하는 가상 NFC 연결 시스템으로 연결함으로써 개방형 무선 인터페이스 플랫폼으로 자리매김하고 있다.

NFC 기술은 다양한 근거리 무선통신들의 다양한 성질들을 결합하여 단말기의 ON/OFF와 관계없이 항상 리더기를 통해 인식이 가능한 카드 에뮬레이션 모드, NFC 활성화 상태에서 RFID Tag 정보를 인식하여 휴대폰이 카드 리더기로 작동하는 리더/라이터 모드, 두 대의 NFC 휴대폰이 카드 리더기로 작동하여 데이터를 상호 간에 전송할 수 있는 P2P 모드로 작동 가능하다. 이 모드들은 RFID와 같이 데이터를 읽고 수정할 수 있는 모드, 블루투스 등과 연결하여 데이터 통신 서비스 가능 모드, 카드에 탑재되어 비접촉식 카드의 성질을 지원하면서 스마트 카드의 안정성을 기반으로 하는 안전한 서비스들이 지원 가능하다.

## 1-2 안드로이드

안드로이드는 모바일 디바이스들을 위한 소프트웨어 스택(Stack)으로써, 운영 체제(OS) 및 미들웨어(Middleware), 그리고 핵심 애플리케이션(Key Application)을 포함하고 있다. 또한 안드로이드는 특정 Vender에 종속되지 않고 개방되어 있어 무료라는 원칙을 가지고 있는 모바일 플랫폼이다. 이러한 오픈 플랫폼인 안드로이드의 대표적인 특징 5가지는 다음과 같다.

### ① 리눅스 커널 2.6을 사용한다.

리눅스(Linux) 커널을 이용하여 운영 체제가 구성되어 있으므로 안드로이드 플랫폼 또한 오픈 소스인 것이다. 대신에, 기존의 리눅스 커널은 GPL 라이선스를 사용하기 때문에 실제 Vender들이 사용하기에는 많은 제약 사항이 있었다. 하지만 안드로이드에서는 커널 등 일부를 제외한 나머지는 LGPL 라이선스보다도 낮은 BSD나 Apache 라이선스를 사용하기 때문에, Vender들이 상용화하기 쉽게 구성하였다는 장점이 있다.

### ② Java 언어의 사용

내부 커널은 리눅스이므로 역시 C언어를 사용하고 있고, 라이브러리의 경우는 C/C++을 사용하고 있으나 애플리케이션 프레임워크 부분부터는 Java를 사용하고 있기 때문에 안드로이드만의 전용 JVM인 Dalvik 가상머신을 통해서 애플리케이션을 구동하고 있다. Java 언어와 C 언어 사이는 JNI(Java Native Interface)를 이용해 연결하여 사용하고 있는데 안드로이드에서는 이를 NDK라는 형태로 자동화하여, 개발자가 사용할 수 있도록 지원하고 있다.

### ③ 다양한 라이브러리의 제공

안드로이드에서는 다양한 라이브러리를 포함하고 있으며, 이를 API로 제공하고 있다. OpenGL ES를 통해서 3D를 지원하고 있고 Webkit을 통해 웹 브라우저를 지원하며, FreeType을 통해서 폰트를 지원하고 있다. 이 외에도 다양한 센서 및 통신에 관한 라이브러리와 디바이스 드라이버들을 지원하고 있다. 여기서 안드로이드의 큰 특징 중 하나는 애플리케이션이 직접 하드웨어와 통신하는 것이 아니라 Binder를 통해 하드웨어와 통신함으로써 응용프로그램이 직접 Root 권한을 얻지 않아도 되고, blocking 방식에서 문제가 발생해도 복구하기가 용이하다.

### ④ Eclipse에 기반한 개발 환경 지원

Eclipse에 기반한 통합 개발환경을 지원한다. 특히 x86에서도 QEMU 기반의 ARM 에뮬레이터를 통하여, 실제 폰과 같이 환경을 제공해주고 있다. 또한, 디버깅 및 메모리 및 성능 프로파일링을 지원하여, 애플리케이션 개발자가 손쉽게 개발할 수 있는 환경을 제공하고 있다.

### ⑤ Bionic C 라이브러리 사용

기존 리눅스와 같이 glibc를 사용하는 것이 아닌 Bionic C 라이브러리를 사용하고 있다. 따라서 기존 리눅스의 glibc를 이용한 소스를 그대로 컴파일(Compile)하기에는 무리가 있어 호환성을 유지하고는 있기는 하나, 약간의 수정을 거쳐야 하는 경우가 있다.

## 1-3 자바

자바는 썬 마이크로시스템즈의 제임스 고슬링(James Gosling)과 다른 연구원들이 개발한 객체 지향적 프로그래밍 언어이며, 썬 마이크로시스템즈(Sun Microsystems)에서 무료로 제공하고 있다. 자바의 개발자들은 유닉스 기반의 배경을 가지고 있었기 때문에 문법적인 특성은 파스칼이 아닌 C++의 조상인 C 언어와 비슷하다. 자바를 다른 컴파일 언어와 구분짓는 가장 큰 특징은 컴파일된 코드가 독립적인 플랫폼이라는 점이다. 자바 컴파일러는 자바 언어로 작성된 프로그램을 바이트코드라는 특수한 바이너리 형태로 변환한다. 바이트코드를

실행하기 위해서는 JVM(자바 가상머신; Java Virtual Machine)이라는 특수한 가상머신이 필요한데, 이 가상머신은 자바 바이트코드를 어느 플랫폼에서나 동일한 형태로 실행시킨다. 때문에 자바로 개발된 프로그램은 CPU나 운영 체제의 종류에 관계없이 JVM을 설치할 수 있는 시스템에서는 어디서나 실행할 수 있으며, 이 점이 웹 애플리케이션의 특성과 맞아떨어져 폭발적인 인기를 끌게 되었다.

#### 1-4 PHP

PHP는 웹 사이트 제작을 비롯한 웹 응용 프로그램을 개발하는 웹 프로그래밍 언어로, 일반적으로 HTML/CSS 및 자바 스크립트와 함께 사용된다. HTML/CSS는 포토샵을 비롯한 다양한 응용 프로그램으로 작성된 이미지, 동영상, 사운드, 텍스트 등을 웹 페이지에 표현하는 스크립트 언어이다. HTML/CSS 코드로만 작성된 웹 페이지는 사용자 게시판, 쇼핑몰 장바구니, 배송 조회처럼 사용자와 상호작용하는 페이지(동적 웹 페이지)도 있기 때문에 HTML/CSS 외에 PHP와 같은 웹 프로그래밍 언어가 필요하다.

PHP는 C 언어와 유사한 점이 많기 때문에 C 언어에 익숙한 사람이라면 누구나 쉽게 익힐 수 있다. PHP는 윈도우에서만 동작하는 ASP.NET과 달리 윈도우는 물론이고 리눅스 같은 유닉스 계열의 운영 체제에서도 사용할 수 있다. 그렇기 때문에 리눅스에서 구축한 웹 사이트의 서버를 윈도우 서버로 변경하더라도 코드 수정이 적다. 데이터베이스도 ASP가 Microsoft의 MSSQL만 사용할 수 있는 반면, PHP는 공개용 데이터베이스인 MySQL뿐만 아니라 Oracle, PostgreSQL, Sybase 등 다양한 데이터베이스를 지원한다. PHP는 간편하게 설치할 수 있고 배우기도 쉬워 JSP보다 개발 기간이 짧다. 또한 내부 엔진이 가벼워 웹 사이트 운영 시 처리 속도가 빠르며, 리눅스와 MySQL 데이터베이스가 무료라 비용 측면에서도 유리하기 때문에 중소 규모의 웹 사이트 개발에 많이 이용된다.

#### 1-5 MySQL

MySQL은 세계적으로 가장 널리 사용되고 있는 오픈소스 데이터베이스로 C/C++, Java



등과 연결 가능한 API도 제공되며, 다양한 웹서버와도 연결이 용이하다. 이러한 오픈소스 데이터베이스인 MySQL의 특징은 다음과 같다.

#### ① MySQL은 관계형 데이터베이스 관리 시스템이다.

관계형 데이터베이스는 데이터를 하나의 커다란 저장 공간에 저장하지 않고 서로 별개의 테이블에 나누어서 저장을 하는 시스템으로 이를 통해 처리 속도와 유연성이 확보된다. SQL(Structured Query Language)은 ANSI/ISO 표준에서 정의한 데이터베이스 접속을 위한 가장 일반적인 표준 언어이다.

#### ② MySQL 소프트웨어는 오픈 소스이다.

MySQL 데이터베이스는 GPL(GNU Public License)을 준수하는 오픈소스 데이터베이스이며, GPL을 준수해서 사용하는 모든 사용자에게 무료로 배포되고 있다.

#### ③ MySQL 서버는 클라이언트/서버 또는 임베디드 시스템에서 사용할 수 있다.

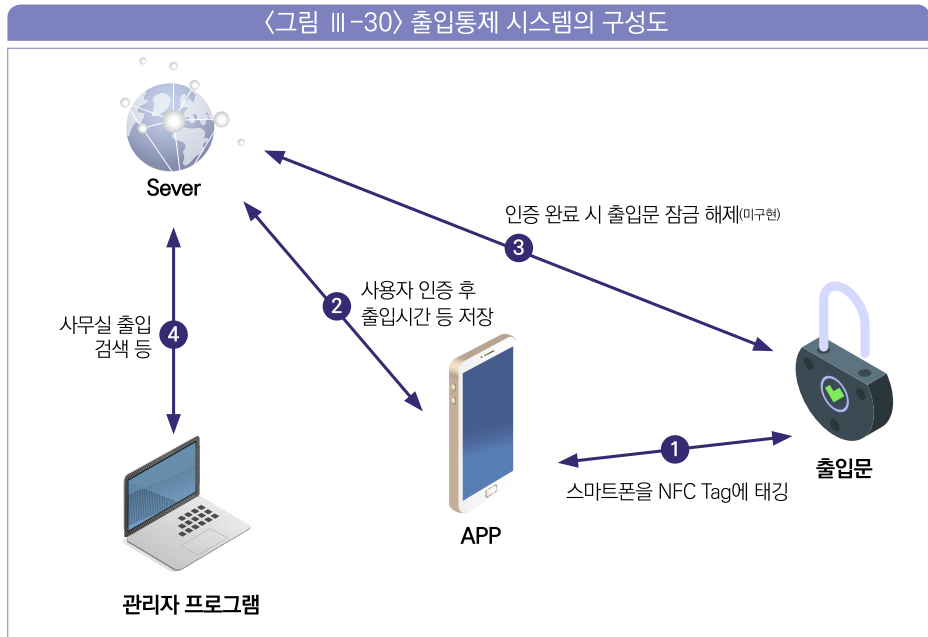
MySQL 데이터베이스 소프트웨어는 다중-쓰레드 SQL 서버로 구성된 클라이언트/서버 시스템이다. 또한, MySQL 서버를 임베디드 형태로도 다양하게 사용할 수가 있다.

### 1-6 NFC를 활용한 출입통제 시스템 구조

NFC를 활용한 출입통제 시스템은 NFC 기능을 장착한 스마트폰의 R/W 기능과 NFC Tag를 이용하였고 구현하기 위하여 안드로이드, Apache, PHP, MySQL, 자바를 이용한다. 스마트폰 애플리케이션은 관리자용과 사용자용 두가지를 구현하였다. 관리자 애플리케이션은 NFC Tag에 사무실 정보를 입력하는 데 사용되고 사용자 애플리케이션은 NFC Tag에서 받은 출입문 정보와 함께 사용자 정보, 현재 시간을 출입통제 서버로 전송한다.

출입통제 서버는 사용자 애플리케이션으로부터 받은 정보를 MySQL과의 질의를 통해 인증 여부를 확인하여 출입승인 여부를 결정한다. NFC를 활용한 출입통제 시스템은 출입문 개폐장치와 직접 연동하지는 않았지만 출입승인정보를 서버에 XML 파일로 저장하고 사용자 스마트폰에선 XML파일을 읽어와 출입승인 여부를 표시하는 형태로 구현한다. 자바

프로그램은 관리자용 프로그램으로써 관리자가 사용자 정보관리 및 출입조회를 쉽게 할 수 있도록 구현한다.



출처: 중부대학교(2014), "NFC를 활용한 출입 통제 시스템 (Access Control System Using NFC)"

## 2 AR, VR, 메타버스와 결합 서비스 소개

### 2-1 사물인터넷과 AR, VR

삼성과 구글, 마이크로소프트, HP 등의 글로벌 ICT 기업들 또한 VR/AR 기술이 스마트폰의 뒤를 이을 미래 신산업으로 성장할 것으로 예상하고 기술 개발과 투자를 확대하며 시장 규모를 키워 가고 있다. 이와 같은 밝은 전망 속에 조사기관별 차이는 있지만 2022년 글로벌 VR/AR 분야의 시장 규모는 약 1,000억 달러에 도달할 것으로 예상되어 향후 산업·경제적 파급효과가 엄청날 것으로 전망된다.

현실을 사실적으로 체험할 수 있는 몰입형 3D 콘텐츠 기반의 VR/AR 서비스에 대한 대중들의 반응도 폭발적이다. 실례로 평창동계올림픽 ICT 체험관에 설치된 VR 미니어처 롤러코스터는 관람객이 실제 롤러코스터를 탄 것과 같은 짜릿함과 스릴감을 느낄 수 있어 큰 인기를 끌었다. 또한, 평창 알펜시아 슬라이딩 센터와 정선 알파인 경기장을 그대로 구현한 VR 콘텐츠를 훈련에 활용한 우리나라 루지팀과 미국 스키, 스노우보드팀은 실전과 같은 훈련을 통해 높은 경기력을 선보여 좋은 성과를 거두었다.

〈그림 III-31〉 VR 훈련 콘텐츠를 활용한 미국 스키팀



출처: U.S. Ski & Snowboard(2018), Winter Olympics : US skiers use VR goggles & brain zapping headphones, 2018. 2.

산업 분야에서는 생산성 향상을 위해 VR/AR 기술을 적극 도입하고 있다. VR/AR, CAD, 제품수명주기관리(PLM; Product Lifecycle Management) 등 다양한 소프트웨어를 제작하는 기업 PTC는 씽웍스(ThingWorx)라는 IoT 개발 플랫폼과 AR이 결합된 소프트웨어를 공개하였다. 씽웍스를 사용하여 사용자는 IoT 제품의 3D 설계 모델을 AR로 가시화하거나, IoT 장치가 수집한 데이터를 시각화하는 AR 애플리케이션을 개발할 수 있다.

프랑스의 에너지 관리 및 자동화 전문기업 슈나이더 일렉트릭은 산업 현장의 스마트 공장화를 위해 산업용 사물인터넷(IIoT; Industrial IoT) 기반 산업 솔루션에 AR 기술을 접목하였다. IIoT는 건물, 공장 설비, 차량, 창고, 팔레트(화물 운반 받침대), 작업자, 로봇 등을 모두 네트워크로 연결시켜 제조 공정 및 에너지 관리 효율성과 생산성을 향상시키고 최적화하는 데 도움을 줄 수 있다.

이에 최근 제조업의 경쟁력을 높일 핵심 기술로 각광받고 있다. 제조 설비 기기와 제조 공정에 AR이 적용된 IIoT 솔루션은 사용자의 단말을 통해 공정 데이터를 실시간으로 모니터링할 수 있다. 또한, 필요한 매뉴얼이나 도면 등을 열람하거나, 기계 내부 상태를 가상 환경에서 살펴볼 수 있는 기능을 갖춰 효율적인 유지 보수가 가능하다. 아래 <그림 III-32>은 슈나이더 일렉트릭의 AR 솔루션을 도입하여 가상의 디스플레이를 통해 제조 설비의 실시간 값을 모니터링하는 시연 장면을 보여주고 있다. 유사 사례로서 공장 자동화 솔루션 전문기업 로크웰 오토메이션(Rockwell Automation)은 작업자가 마이크로소프트의 홀로렌즈(HoloLens) 헤드셋을 착용하고 장비를 바라보면 관련 정보가 가시화되어 업무 효율성을 향상시킬 수 있는 솔루션을 공개했다.

<그림 III-32> 슈나이더 일렉트릭의 AR 기반 IIoT 솔루션을 도입한 흥아기연사의 카톤 포장기



출처: 흥아기연

## 2-2 사물인터넷과 메타버스

오늘날 융합산업은 IoT를 활용한 디지털 트윈에 초점을 두고 있다. 전통산업 분야에서 자율형 IoT를 활용한 융합산업의 유지 관리 및 안정성에 대한 정보를 제공함으로써 전통제품이 보다 효과적으로 사용될 수 있도록 지원할 수 있다.

이처럼 가상세계의 메타버스에 의해 구현된 융합산업을 위한 디지털 트윈은 아래 <그림 Ⅲ-33>과 같이 자율적인 트윈(Autonomous Twin)로의 발전하게 될 것이다. 자율트윈의 상호연동은 전통산업의 기계 장치와 공장 설비와 유사한 가상세계를 구축함으로써 초연결(Linked internet)에 의해 메타버스 기반의 지능형 트윈이 융합산업의 모델이 될 것이다.

이러한 메타버스 자율트윈이 동기화되고 지능화되면서 집단지성 기반의 연결형 트윈은 사회 연계형(Socialization)으로 다양한 전통산업의 비대면 고객들을 대상으로 글로벌 연결형 'AutoTwin Internet'이 융합산업의 핵심요소로서 중심이 될 것이다.



출처: 한국인터넷진흥원(2021), “비대면 융합산업 발전과 디지털 대전환”



## 참고문헌

- 윤성진(2021), 딥러닝 교과서, 이지스 퍼블리싱
- 브래들리 에프론, 헤스티 (2019), 컴퓨터 시대의 통계적 추론(이병욱 역)
- 빅데이터 비즈니스 성공지도(2014), 에이콘
- 박성수(2020), 강화학습 원리와 알고리즘, 위키 북스
- 니콜라스 카(2008), 빅스위치, 동아시아
- 강민구(2021), 비대면 융합산업 발전과 디지털 대전환, KISA
- 최동진(2019), 5G 시대의 차세대 IoT 보안, 정보통신기획평가원
- 나영식(2020), 제조용 IoT, KISTEP 한국과학기술기획평가원
- KOTRA(2021), 전기전자산업 해외 시장동향 및 시사점
- 최진철, 손영상(2018), IoT 및 AI 융합을 통한 VR/AR 발전 전망, ITFIND  
정보통신기술진흥센터
- 김영상, 유성민(2017), 4차 산업혁명과 IoT-AI 플랫폼, 한국정보기술학회지
- 삼정KPMG 경제연구원(2017), 금융산업 4차 산업혁명과 만나다
- 중부대학교 정보보호학과(2014), NFC를 활용한 출입통제 시스템
- 황종모(2016), 금융권 사물인터넷(IoT) 동향 및 향후 전망, 전자금융과 금융보안
- 티모시 추(2020), IoT 비즈니스 전략, 길벗
- 최환석, 이우섭(2015), 사물인터넷 플랫폼 기술 및 국제 표준화 동향, 방송과 미디어
- 조동군, 정우수, 박수우(2014), 사물인터넷(IoT) 도입효과 분석 연구,  
한국정보통신진흥협회
- ETRI(2019), 스마트 IT융합 플랫폼을 위한 지능형 센서 기술 동향
- Maxim Integrated(2018), 미래의 산업용 IoT 실현을 위해 필요한 것은?
- 과기정통부(2019), 5G기반스마트공장고도화전략(안)
- 김영식(2015), 사물인터넷 시대의 사이버 물리 시스템 보안 기술 동향

- 금융보안원(2016), 금융권 사물인터넷(IoT) 동향 및 향후 전망
- KISTEP 한국과학기술기획평가원(2020), KISTEP 기술동향브리프
- 한국인터넷진흥원(2021), "비대면 융합산업 발전과 디지털 대전환"
- 삼정KPMG 경제연구원(2018), 4차 산업혁명과 제조혁신:스마트 팩토리 도입과 제조업 패러다임 변화
- Getty Images Bank, [www.gettyimagesbank.com](http://www.gettyimagesbank.com)
- The HDF group, [www.hdfgroup.org](http://www.hdfgroup.org)
- STHDA, <http://www.sthda.com>
- ITU, [www.itu.int](http://www.itu.int)
- 위키피디아, [www.wikipedia.org](http://www.wikipedia.org)
- 구글 클라우드, [www.cloud.google.com](http://www.cloud.google.com)
- 아마존 AWS, [aws.amazon.com](http://aws.amazon.com)
- MS Azure, [azure.microsoft.com](http://azure.microsoft.com)
- Amazon 2016 Annual Report, [www.annualreports.com](http://www.annualreports.com)
- 베스핀글로벌, [www.bespinglobal.com](http://www.bespinglobal.com)
- 메가존, [www.megazone.com](http://www.megazone.com)
- 쇼피파이, [www.shopify.com](http://www.shopify.com)
- 애플, [www.apple.com](http://www.apple.com)
- Youtube(2018), "Why HSBC moved its more than 100 petabytes of data to Google Cloud"
- BBVA, [www.bbva.com](http://www.bbva.com)
- AlleyWatch, [alleywatch.com](http://alleywatch.com)
- 디지털이니셔티브그룹, [www.digitaltransformation.co.kr](http://www.digitaltransformation.co.kr)
- 넷플릭스, [www.netflix.com](http://www.netflix.com)



- 유튜브 (2021), Learn with Shopify, "The Official Shopify Tutorial for 2021: Set Up Your Store the Right Way"
- 한국인터넷진흥원, [www.kisa.or.kr](http://www.kisa.or.kr)
- GS칼텍스 미디어 허브, [gscaltexmediahub.com](http://gscaltexmediahub.com)
- 한국정보기술학회지, [ki-it.or.kr](http://ki-it.or.kr)
- LG CNS, [blog.lgcns.com](http://blog.lgcns.com)
- 벤처스퀘어, [www.venturesquare.net](http://www.venturesquare.net)
- SK텔레콤, [www.sktelecom.com](http://www.sktelecom.com)
- Marketandmarkets, [www.marketsandmarkets.com](http://www.marketsandmarkets.com)
- Imperva, [www.imperva.com](http://www.imperva.com)
- Rohm Semiconductor, [www.rohm.co.kr](http://www.rohm.co.kr)
- 포스코경영연구원, [www.posri.re.kr](http://www.posri.re.kr)
- 한국반도체산업협회, [www.kisa.or.kr](http://www.kisa.or.kr)
- BCC 리서치, [www.bccresearch.com](http://www.bccresearch.com)
- 바이탈리티, [www.vitality.co.uk](http://www.vitality.co.uk)
- 에너지관리공단, [www.energy.or.kr](http://www.energy.or.kr)
- 노션, [www.getnotion.com](http://www.getnotion.com)
- 삼성전자, [www.samsung.com](http://www.samsung.com)
- 프로그레시브, [www.progressive.com](http://www.progressive.com)
- GMS, [www.global-mobility-service.com](http://www.global-mobility-service.com)
- Kawasaki, [www.kawasaki.com](http://www.kawasaki.com)
- U.S. Ski & Snowboard(2018), Winter Olympics : US skiers use VR goggles & brain zapping headphones, 2018. 2
- 스마트제조혁신추진단, [www.smart-factory.kr](http://www.smart-factory.kr)



- 아우디, [www.audi.co.kr](http://www.audi.co.kr)
- 벤츠, [www.mercedes-benz.co.kr](http://www.mercedes-benz.co.kr)
- 보쉬, [www.bosch.co.kr](http://www.bosch.co.kr)
- ST마이크로일렉트로닉스, [www.st.com](http://www.st.com)

## 교재 집필 위원

### 이병욱 | 제1장

- 한국과학기술원(KAIST) 전산학과 석사
- (전) BNP Paribas Cardif 전무
- (전) 삼성생명 주식회사 수석
- (현) 서울과학종합대학원 디지털금융 주임교수
- (현) 인공지능연구원(AIRI) 부사장
- (현) 한국금융연수원 겸임교수
- **저서** <블록체인 해설서>(에이콘, 2019)  
<비트코인과 블록체인, 가상자산의 실제>(에이콘, 2020)  
<돈의 정체>(에이콘, 2021)

### 정명훈 | 제2장

- 서울과학종합대학교 경영대학원 석사
- (전) 한국오라클 부장
- (현) 구글코리아 이사
- (현) 경북대 겸임교수
- (현) 정보통신산업진흥원 평가위원

### 이강주 | 제3장

- 시카고대학교 경영대학원 석사
- (전) 모닝스타 데이터 애널리스트
- (전) H&Consulting 팀장
- (전) 지바이크 CEO
- (현) 스파크랩 이사
- **역서** <IoT 비즈니스 전략: 5가지 프레임워크로 배우는 사물인터넷 원칙과 솔루션>(길벗, 2020)

## 감수자

### 장현기

- 서울대학교 물리학과 박사
- (전) 삼성전자 책임
- (전) IBM 실장
- (전) SK Holdings 팀장
- (전) 신한은행 디지털 R&D센터 본부장
- (현) 신한금융지주 디지털 플랫폼 추진단 단장

### 신용녀

- 고려대학교 컴퓨터학과 박사
- (전) 한국인터넷진흥원 책임연구원
- (전) 한국은행 전자금융팀 과장
- (전) 한양사이버대학교 컴퓨터공학과 교수
- (전) 아마존웹서비스(AWS) 기술이사
- (현) 한국마이크로소프트 NTO(최고기술임원)

### 한규영

- 경북대학교 전자공학과 학사
- 대구대학교 경영대학원 석사
- (현) SK텔레콤 부장

# 헬로, 핀테크!

핀테크 기반기술

초판 1쇄 2021년 11월 29일  
집필위원 이병욱, 정명훈, 이강주  
감수자 장현기, 신용녀, 한규영  
발행인 변영한  
발행처 사단법인 한국핀테크 지원센터  
(04213) 서울시 마포구 마포대로 122, 11층, 12층  
전화 02) 6375-1550  
홈페이지 [www.fintech.or.kr](http://www.fintech.or.kr)  
등록 2020년 7월 24일(제879-82-00208호)  
I S B N 979-11-92068-10-7 05320  
979-11-92068-02-2 05320(세트)

※ 본 도서는 한국핀테크지원센터의 허가 없이 무단 전재 또는 복사를 금하며,  
적발 시 저작권법에 의하여 민·형사상의 책임 및 징역·벌금 등의 불이익을 당할 수 있습니다.

「헬로, 핀테크」도서의  
보조학습자료로서 동영상 강의를  
FinEDU에서 제공하고 있습니다.



[finedu.fintech.or.kr](http://finedu.fintech.or.kr)

「헬로, 핀테크」시리즈 외에도  
한국핀테크지원센터에서  
엄선하여 기획한 핀테크 전문 커리큘럼  
FinEDU 코스를 경험하세요!



### 헬로, 핀테크! 입문

핀테크 개요, 기술, 시장, 핀테크 관련 법률,  
금융회사와 핀테크 기업의 협업 등  
한국핀테크지원센터 지음



### 헬로, 핀테크! 지급결제·송금

핀테크 지급결제·송금 기술, 규제 및 정책 동향,  
시장 및 산업 동향 등  
한국핀테크지원센터 지음



### 헬로, 핀테크! 금융플랫폼·금융데이터

금융정보플랫폼, 오픈뱅킹플랫폼, P2P플랫폼,  
금융빅데이터 기술 및 활용사례, 마이데이터 산업 등  
한국핀테크지원센터 지음



### 헬로, 핀테크! 자산관리·보험

자산관리테크 및 인슈어테크 서비스,  
기반 기술, 규제 및 정책 등  
한국핀테크지원센터 지음



### 헬로, 핀테크! 보안인증·블록체인

보안인증 핵심기술 및 사례, 블록체인 기술,  
정책 및 산업 동향, 레그테크 기술 및 사례 등  
한국핀테크지원센터 지음



### 헬로, 핀테크! 개인신용정보 관리 및 활용

개인신용정보 관련 법규, 개인신용정보 수집 및 관리  
실무, 개인신용정보 기술현황, 시장 및 산업 동향 등  
한국핀테크지원센터 지음



### 헬로, 핀테크! 핀테크 기반기술

빅데이터와 인공지능 기술, 클라우드 서비스 등  
한국핀테크지원센터 지음

# HELLO, FINTECH



금융위원회



한국핀테크지원센터  
Fintech Center Korea



한국금융연수원  
KOREA BANKING INSTITUTE