

03

ICT 요소기술 개론

VI. 정보기술 기본 128

- 14. 정보기술 기초
- 15. 빅데이터의 이해
- 16. 빅데이터 분석 프로세스




VII. 통신기술 기본 167

- 17. 정보통신 기초
- 18. 정보통신망

VIII. 네트워크 기본 197

- 19. 네트워크 기초
- 20. 네트워크 구조

VI. 정보기술 기본

경북대학교 김상현 교수   

14. 정보기술 기초

14.1. 컴퓨터 시스템 구성

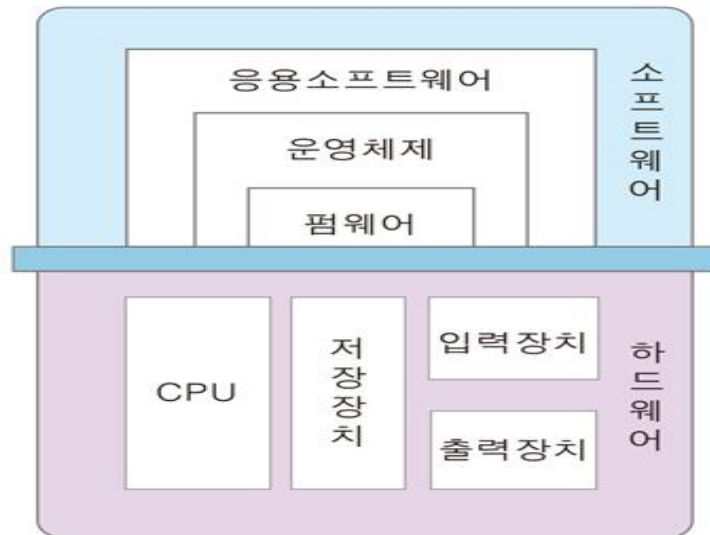
컴퓨터 시스템은 1946년 등장한 최초의 전자식 컴퓨터 애니악(ENIAC)에서 시작하여 폰노이만(von Neumann) 컴퓨터로 불리는 프로그램 내장방식의 시스템이 개발된 이래 현재까지 지속해서 발전해 왔다. 일반적으로 컴퓨터 시스템은 크게 이를 구성하는 하드웨어와 하드웨어를 작동시키기 위한 소프트웨어의 조합으로 구성된다.

하드웨어(hardware)는 눈에 보이고 만질 수 있는 외형을 가진 전기적, 기계적인 장치로서 중앙처리장치(CPU, Central Processing Unit), 주기억장치, 입력장치, 출력장치로 구성되어 있다. 중앙처리장치는 전체 컴퓨터 시스템을 제어하고, 시스템 연산을 처리하는 역할을 하는 장치이다. 모든 작동 과정이 CPU의 제어를 받기 때문에 컴퓨터의 두뇌에 해당하며 연산장치, 제어장치, 레지스터로 구성된다. 대형 컴퓨터는 중앙처리장치가 보드 형태로 제작이 되고 소형 컴퓨터에서는 마이크로프로세서 칩의 형태로 제작된다.

주기억장치(main memory)는 컴퓨터에 필요한 정보와 프로그램을 저장하는 공간으로 주기억장치와 보조기억장치로 구성된다. 주기억장치는 CPU에 의해 즉시 처리할 명령어나 데이터를 저장하는 역할을 하는 것으로 데이터를 읽거나 쓰기 위해서 사용하는 메모리 공간이다. 전원이 꺼지면 내용이 모두 지워지는 휘발성 메모리인 RAM(Random Access Memory)과 전원이 끊어져도 기억된 내용이 보존되는 ROM(Read Only Memory)이 여기에 해당한다. 보조기억장치는 주기억장치를 보조하는 역할을 하는 것으로 주기억장치가 속도가 빠르지만, 용량이 적다는 단점을 보완한다. 보조기억장치로 주로 사용되는 메모리인 하드디스크, 광학 디스크(블루레이 디스크, CD-ROM, DVD), 플래시메모리 등이 있다.

입력장치(input device)는 컴퓨터 시스템 외부로부터 그림, 문자, 소리, 영상 등의 데이터를 입력받는 장치로 키보드, 마우스, 터치패드, 터치스크린, 카메라, 마이크, 바코드 리더기, 광학 스캐너 등이 있다. 대형

컴퓨터의 형태에서 소형의 모바일 기기들이 빠르게 발전하고 사용 비중이 확대되면서 터치패드, 햅틱 인터페이스 등을 사용한 입력장치들이 더 주목받고 있다.



[그림 3-1] 컴퓨터 시스템의 체계

출력장치(output device)는 중앙처리장치에서 처리한 결과물을 컴퓨터 시스템 외부로 전달하는 역할을 하는 장치로 인쇄, 음향, 영상 등 다양한 매체를 통해 사용자에게 정보를 제공한다. 출력장치로는 프린터, 스피커, 모니터 등이 있다. 컴퓨터 기술과 멀티미디어 기술이 발전하면서 멀티미디어 정보가 많은 비중을 차지하고 있으며, 이를 위해 다양한 촉각을 활용한 출력장치들이 개발되고 있다.



[그림 3-2] 소프트웨어와 언어처리

컴퓨터 시스템의 하드웨어는 스스로 작동하는 것이 아니라 이를 작동시키기 위한 프로그램이 있어야 한다. 이것이 바로 소프트웨어(software)로 컴퓨터 시스템의 하드웨어를 작동시켜 원하는 작업 결과를 얻기 위해 실행되는 프로그램이나 명령어들을 의미한다. 소프트웨어에는 크게 시스템 소프트웨어와 응용 소프트웨어가 있다. 먼저, 시스템 소프트웨어(system software)는 응용 소프트웨어가 실행될 때 하드웨어를 효율적으로 사용하도록 메모리, CPU 등의 자원을 배치하고 관리해

주는 기능을 수행하는 역할을 한다. 시스템 소프트웨어에는 하드웨어를 작동시키고 자원을 관리해주는 운영체제, 대용량의 데이터를 관리하기 위한 데이터베이스, 응용 프로그램을 개발하기 위한 컴파일러, 네트워크 접속프로그램, 프린터 드라이버 등의 유틸리티 소프트웨어가 있다.

응용 소프트웨어(application software)는 사용자가 요구하는 특정 업무를 처리하기 위해 작성된 프로그램으로 스프레드시트, 프레젠테이션 등의 개인 사무용 소프트웨어, 데이터 시각화, 통계분석 등 과학분석용 소프트웨어, ERP, SCM, CRM, 전자상거래 등 비즈니스 활용을 위한 기업용 소프트웨어 등이 있다.

컴퓨터 시스템이 작동되는 원리를 살펴보면 컴퓨터에서 전원을 켜올 때 시스템 내 하드웨어를 점검하고 운영체제가 작동할 수 있도록 초기화 작업을 수행하게 된다. 이런 과정은 모두 ROM 메모리에 기록되어 있으므로 전원을 처음 켜올 때 컴퓨터 시스템은 ROM으로부터 시작 프로그램을 읽고 해당하는 소프트웨어를 실행시킨다.

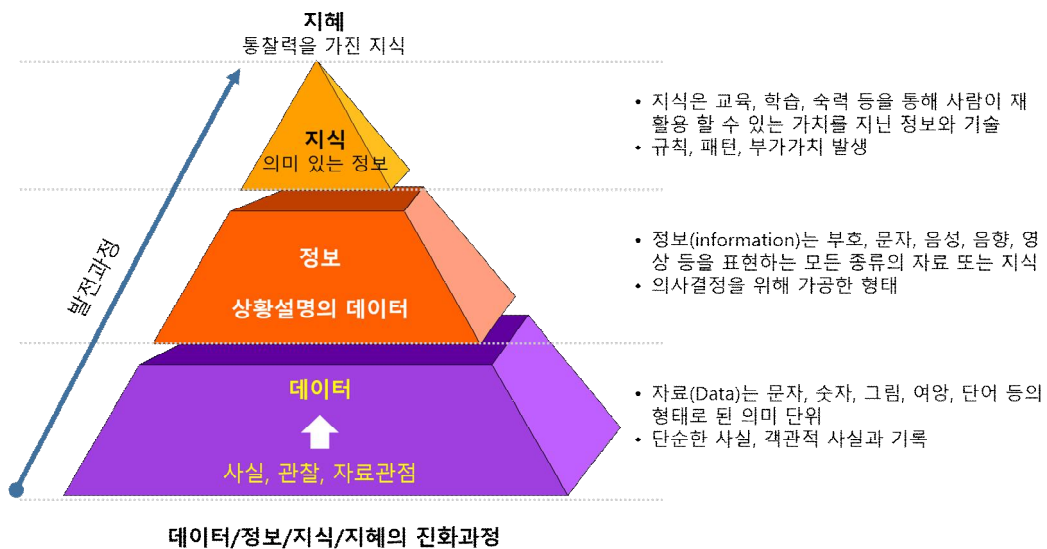
이때 컴퓨터를 작동시키기 위해 실행되는 소프트웨어를 펌웨어(firmware)라 한다. 펌웨어는 하드웨어를 초기화하여 사용 가능한 상태로 만들고 운영체제를 RAM으로 가져와서 CPU가 운영체제를 실행할 수 있도록 준비한다. 운영체제가 주기억장치인 RAM에 로드되면 운영체제는 프로세서 관리, 디바이스 관리, 주기억장치 관리, 보조기억장치 관리, 사용자 인터페이스 관리, 응용 프로그램 관리 등을 통해 전체 컴퓨터 시스템을 관리한다.

14.2. 정보의 개념

컴퓨터의 등장으로 사회가 빠르게 발전하면서 처리해야 할 정보의 양이 빠른 속도로 늘어나게 되고 처리해야 하는 양도 기하급수적으로 증가하고 있다. 이에 컴퓨터 시스템 역시 더 빠르게 발전되어 왔으며 정보가 중요한 역할을 차지하는 사회를 형성하게 되었다. 사람들에게는 수시로 접하는 자료와 정보들을 선별해서 관리하고 활용하는 것이 매우 중요하게 되었다. 자료와 정보라는 개념을 혼동해서 사용하는 경우가 많지만, 이 두 가지의 개념에는 차이가 있다. 먼저 자료(data)는 단순한 관찰이나 측정을 통해서 얻게 되는 값이나 사실 등을 의미한다. 일상생활에서 쉽게 접할 수 있는 단순한 수치나 지식과 같이 가공되지 않은 상태의 것으로 그 자체만으로는 의미가 없다.

정보(information)는 수많은 자료 중에서 필요한 것을 선별하여 가공 처리해서 유용하고 가치 있는 형태로 정리한 결과를 의미한다. 자료는 컴퓨터에 입력된 상태로 구조화되지 않은 원재료에 해당하며, 이를 선별하고 정리하고 가공·처리하여 유용한 형태로 변환시킨 결과가 정보에 해당한다. 예를 들어 날씨가 궁금한 사용자에게 일기예보 부분은 정보가 될 수 있지만, 일기예보의 내용이 들어있는 신문 자체는 단순한 자료일 수 있다.

환율이나 주식시세와 같은 내용도 사용자에게 따라서 자료가 될 수도 있고 정보가 될 수도 있다. 즉, 자료와 정보는 사용하는 방법이나 용도, 수용자에 따라 자료가 될 수도 있고 정보도 될 수도 있으며 정보들이 모여서 새로운 정보를 만들어낼 수도 있다. 더 나아가 지식(knowledge)은 정보를 알고 이를 활용할 수 있는 것을 의미하며 이 지식을 적절한 시간과 장소에서 유용하고 가치 있게 활용할 수 있는 것을 지혜(wisdom)라고 할 수 있다.



[그림 3-3] 데이터, 정보, 지식 그리고 지혜

정보는 미래 경제를 결정하는 중요한 요소로 사용자가 정보를 어떻게 사용하느냐에 따라 그 가치가 달라질 수 있다. 따라서 정보사회에서는 필요한 정보를 창출하기 위해 가공할 자료와 정보가 요구되고 정보를 바탕으로 의사결정과 행동을 할 수 있는 지식이 수반되어야 한다. 정보가 사용자에게 가치 있고 유용하려면 이를 판단하기 위한 질적 특성을 가져야 하는데 일반적으로 신뢰성, 적시성, 적절성, 이해성, 일관성 등이 있다.

① 신뢰성(reliability)

정보의 신뢰성은 정보 사용자가 정보를 신뢰할 수 있어야 한다는 것으로 이를 확보하기 위해 정보의 사실을 정확하면서 아무런 편견 없이 제공해야 한다. 예를 들어 단순히 사람들의 흥미를 끌기 위해서 부정확한 정보를 제공한다면 정보의 신뢰성을 상실했다고 할 수 있다.

② 적시성(timeliness)

정보 적시성은 사용자가 의사결정을 할 시점에 필요한 정보를 제공해야 한다는 것으로 적시성이 없는 정보는 정보로서의 가치나 유용성을 상실했다고 할 수 있다. 정보의 가치는 일반적으로 생산 시점이 사용 시점에서 멀어질수록 그 효용성이 감소하게 되므로 적절한 시점에 사용자에게 전달되는 것이 무엇보다 중요하다.

③ 적절성(relevance)

정보 적절성은 정보가 사용자의 이용 목적에 적합해야 한다는 것으로 정보를 이용함으로써 정보 사용자의 장래 불확실성을 감소시킬 수 있어야 한다.

④ 이해성(understanding)

정보 이해성은 정보 사용자가 특정 정보에 관해 내용의 의미를 정확하게 이해하는 것을 의미한다.

⑤ 일관성(consistency)

정보 일관성은 일정 기간을 두고 정보를 정기적으로 생산하는 경우 정보 사용자가 정보들을 서로 비교할 수 있어야 한다는 것을 의미한다.

정보의 가치는 크게 규범적 가치, 주관적 가치, 현실적 가치로 구분할 수 있다. 규범적 가치(normative value)는 사람은 언제나 합리적이고 이성적이기 때문에 계산할 때 실수를 하지 않는다고 가정했을 때 정보를 사용할 때와 사용하지 않을 때의 가치를 계산하여 그 차이로 나온 값을 의미한다. 주관적 가치(subjective value)는 개인의 직관에 따라 가치를 판단하는 것을 의미하며 현실적 가치(realistic value)는 실제 시험을 통해 실행해보고 정보를 사용했을 때와 사용하지 않았을 때의 가치를 비교하여 판단하는 것을 의미한다.

정보를 표현할 때에도 여러 가지 유형으로 표현하는데 과거에는 주로 숫자, 문자만을 이용해서 표현했지만, 정보처리 기술이 발전하면서 그림, 동영상, 소리, 애니메이션 등의 다양한 멀티미디어 형태로 표현한다. 정보는 표현하는 형태에 따라 그림 정보, 동영상 정보, 수치 정보, 문자 정보, 소리 정보 등으로 구분하며 정보의 유형에 따라 저장하고 처리, 표현하는 방법이 서로 다르다.

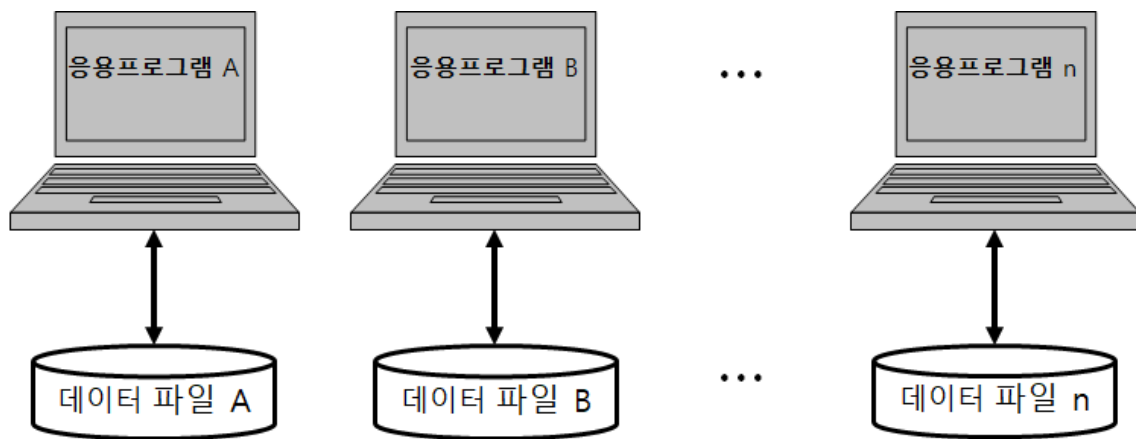
그림 정보는 움직이지 않는 평면 위에 점과 선을 색채를 이용해서 표현하는 정보를 의미하며 말과 글로 표현할 수 없는 부분도 이해하기 쉽게 표현할 수 있다. 동영상 정보는 그림, 소리, 문자를 함께 사용하여 시간에 따라 움직임을 눈과 귀로 보고 들을 수 있는 영상을 의미한다. 수치 정보는 계산이 가능한 숫자 정보로 숫자를 통해 정보의 양이나 크기를 표현할 수 있다. 문자 정보는 한글, 영

문, 특수문자, 숫자 등으로 의사소통을 위해 규정한 기호들로서 글을 표현할 때 사용한다. 소리 정보는 사람의 귀를 통해 들을 수 있는 형태의 정보로 세기나 높낮이의 변화를 통해 감정까지도 전달할 수 있다.

14.3. 데이터베이스 정의

과거에는 데이터를 관리하기 위해 파일 처리 시스템이라는 소프트웨어를 이용했다. 파일 처리 시스템은 데이터를 파일로 관리할 수 있도록 파일을 생성하고 삭제, 수정, 검색하는 기능을 제공하며 운영체제와 함께 설치된다. 응용 프로그램에 따라 필요한 데이터를 별도의 파일로 관리하는 방식으로 데이터 양이 적을 때 적합하고 별도의 구매 비용이 들지 않는다는 장점이 있다.

하지만 응용 프로그램마다 파일을 별도로 유지하고 관리해야 하는 방식 때문에 같은 내용의 데이터가 여러 파일에 중복으로 저장되거나 데이터 파일에 대한 동시 공유, 보안, 회복 기능이 부족하다는 문제점이 있다. 응용 프로그램에 데이터 파일이 종속적이고 응용 프로그램 개발이 쉽지 않다는 문제점도 있다. 이러한 파일 처리 시스템의 문제점과 대량의 데이터가 발생하는 현대 사회에서 파일 처리 방식이 한계가 있다는 점이 지적되면서 이를 보완하기 위한 데이터베이스가 널리 사용되기 시작했다.



[그림 3-4] 파일 처리 시스템

데이터베이스(Database)는 1963년 Development and Management of a Computer-centered Data Base 심포지엄에서 처음 소개된 용어이다. 데이터베이스는 파일 처리 시스템과 다르게 특정 조직의 여러 응용 프로그램이 저장된 데이터를 공유할 수 있도록 데이터를 통합하여 관리한다. 따라서 방대한 데이터를 특정 사용 목적에 맞게 체계적으로 정리해 놓은, 관련 있는 데이터를 모아두는 저장소의 개념으로 볼 수 있다. 일반적으로는 "특정 조직의 여러 사용자가 '공유'하여 사용할 수 있도록 '통합'해서 '저장'한 '운영'

데이터의 집합"이라는 정의를 많이 사용한다. 각각의 핵심 개념을 하나씩 살펴보자.

① 공유 데이터(shared data)

데이터베이스는 특정 조직의 여러 사용자가 함께 소유하고 이용할 수 있어야 하는 데이터이다. 그러므로 사용 목적이 다른 사용자들을 함께 고려해서 데이터베이스를 구성해야 한다.

② 통합 데이터(integrated data)

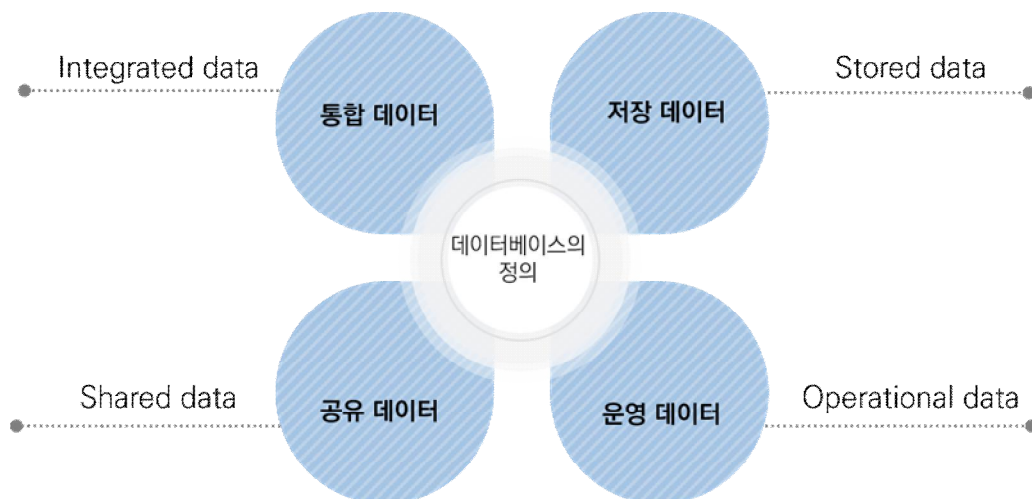
데이터의 중복을 최소화하고 통제가 가능한 최소한의 중복만을 허용하는 데이터이다. 일반적으로 관리하기 어려운 문제가 발생할 수 있으므로 똑같은 데이터가 여러 개 존재하는 중복성을 허용하지는 않지만, 효율성 때문에 의도적으로 중복을 허용하는 때도 있다.

③ 저장 데이터(stored data)

컴퓨터가 접근할 수 있는 매체에 저장되어있는 데이터이다. 데이터베이스의 데이터는 주로 컴퓨터가 처리하기 때문이다.

④ 운영 데이터(operational data)

특정 조직을 운영하고 조직의 주요 기능을 수행하는 데 필요한 데이터이다. 따라서 일시적으로 사용하고 끝나는 것이 아니라 지속해서 유지해야 하는 데이터이다.



[그림 3-5] 데이터와 정보의 처리

데이터베이스는 사용자들의 데이터 요구에 실시간으로 응답할 수 있고, 응용 프로그램에서 실행된 결과에 따라 저장된 데이터도 계속 변화하게 된다. 즉, 현실 세계의 상태를 정확히 반영하여 데이터를 계속 삽입, 삭제, 수정하여 현재의 정확한 데이터를 유지할 수 있게 한다.

여러 사용자가 동시에 이용할 수 있는 동시 공유가 가능한데 여기서 동시 공유라는 것은 여러 사용자가 서로 다른 데이터를 동시에 사용하는 것뿐 아니라 같은 데이터를 동시에 사용하는 것도 모두 지원하는 것을 의미한다. 데이터베이스는 데이터의 값(value)으로 참조할 수 있다. 즉, 저장된 주소나 위치가 아니라 데이터의 내용(content)을 바탕으로 정보를 검색할 수 있는 편의성을 제공한다.

데이터베이스는 데이터베이스 관리 시스템(DBMS, DataBase Management System)의 용어와 혼용되어 사용되기도 하는데 데이터베이스 관리 시스템은 파일 처리 시스템의 데이터 중복과 증속 문제를 해결하기 위해 제시된 소프트웨어를 말한다.

DBMS는 조직에 필요한 데이터를 데이터베이스에 통합하여 저장하고 이에 대한 관리를 집중적으로 담당하며 응용 프로그램을 대신해서 데이터베이스에 있는 데이터를 삽입, 삭제, 수정, 검색하는 작업을 주로 수행한다. 데이터베이스는 이러한 검색과 데이터의 삽입, 삭제, 수정의 빈도에 따라 시스템 구축의 난이도가 결정되는데 그 유형은 크게 네 가지로 나뉜다. 데이터베이스 유형 1의 구축이 가장 쉽고 유형 4의 구축이 가장 어렵다.

[표 3-1] 데이터베이스 구축 난이도에 따른 유형

| 유형 | 검색빈도 | 변경빈도 | 특징 |
|-----|------|------|--|
| 유형1 | 적다 | 적다 | 검색이 많지 않아 데이터베이스를 구축할 필요 없음 보존가치가 있는 경우에 구축 예: 공룡 데이터베이스 |
| 유형2 | 많다 | 적다 | 사용자 수 보통 검색은 많지만, 데이터에 대한 변경은 적음 예: 도서 데이터베이스 |
| 유형3 | 적다 | 많다 | 예약 변경/취소 등 데이터 변경은 많지만, 검색은 적음 실시간 검색 및 변경이 중요함 예: 비행기 예약 데이터베이스 |
| 유형4 | 많다 | 많다 | 사용자 수 많음 검색도 많고 거래로 인한 변경도 많음 예: 증권 데이터베이스 |

(출처 : 데이터베이스 개론과 실습(SQL Server로 배우는), 박우창, 남송휘, 이현룡, 한빛아카데미)

14.4. 데이터베이스 특징

데이터베이스를 사용하기 이전에 파일 처리 시스템은 응용 프로그램마다 별도의 시스템들이 존재하고 필요한 데이터를 각 시스템이 개별적으로 처리하여 사용했다. 따라서 조직 내 다양한 정보를 바탕으로 새로운 정보를 만들기 위해 각 응용 프로그램마다 정보를 수집해야 했으며 이 과정에서 많은 시간과 비용이 들었다.

또한, 정보를 중복하여 입력하거나 잘못된 값을 입력하는 실수들을 응용 프로그램에서 방지하기가 어렵고 동시에 데이터를 공유하여 처리하는 것이 어려웠다. 하지만 데이터베이스의 등장으로 데이터를 관리하던 기존의 방식의 한계점을 보완할 수 있게 되었고 더욱 구조적이고 안전하게 데이터를 관리할 수 있게 되었다. 데이터베이스의 주요 특징을 살펴보면 다음과 같다.

① 데이터 중복성을 최소화할 수 있다.

파일 처리 시스템을 사용하는 경우 응용 프로그램이 조직마다 다를 수 있고 따라서 저장해야 하는 데이터도 달라질 수 있다. 별개의 데이터일 수는 있지만 공통된 데이터가 각각의 응용 프로그램에서 처리될 가능성도 매우 크다. 이는 조직 전체의 관점에서 봤을 때 심각한 데이터 중복성(redundancy)의 문제를 가져올 수 있고 저장 공간의 낭비를 초래할 수도 있으며 전체 데이터의 유지나 보수를 어렵게 만들 수도 있다. 데이터베이스를 이용하면 모든 데이터가 통합되어 하나의 저장소에서 관리되고 업무의 흐름에 따라 데이터를 통합 및 분리하여 관리할 수 있으므로 중복성을 최소화할 수 있다.

② 데이터의 공유성을 최대화할 수 있다.

데이터베이스는 공동 작업에 맞게 구조적으로 설계되고 저장된 값들은 통합된 체계에 따라 유지 및 관리된다. 조직 구성원들은 데이터베이스를 공동으로 소유하고 이용하면서 업무에 필요한 정보를 새롭게 생성하거나 이용할 수 있고 나아가 업무의 효율성도 높일 수 있다.

③ 데이터 일관성을 유지할 수 있다.

데이터베이스는 중복성을 최소화하기 때문에 데이터가 일관성을 유지할 수 있도록 한다. 즉, 데이터 불일치성을 미리 방지하기 때문에 정확한 데이터가 저장, 관리되고 이를 바탕으로 사용자들에게 신뢰할 수 있는 정확한 정보를 제공하여 일관성을 항상 유지하게 된다.

④ 데이터 무결성을 유지할 수 있다.

무결성은 데이터베이스에 정확한 데이터가 유지되고 있음을 보장하는 것으로 인가되지 않은 방법으로

데이터가 변경될 수 없도록 하거나 권한이 있는 사람들에게만 데이터와 정보를 허락하여 그들에 의해서만 수정될 수 있도록 보장하는 것을 의미한다. 데이터베이스는 조건에 맞지 않는 데이터는 입력되지 않도록 하거나 권한이 있는 사람들의 접근만을 허용함으로써 무결성을 유지하게 된다.

⑤ 데이터 독립성을 유지할 수 있다.

독립성은 데이터의 내용과 구조가 응용 프로그램과 독립되어 존재하는 것을 의미한다. 기존의 파일 처리 시스템은 데이터가 각 응용 프로그램마다 별도로 생성되고 프로그램과 밀접하게 연관되어 있어 데이터를 저장할 때 독립성이 확보되지 못했다. 데이터베이스는 이러한 종속성 문제를 해결하여 데이터의 표현방법이나 저장 위치가 변하더라도 응용 프로그램에는 아무런 영향을 미치지 않도록 함으로써 데이터 독립성을 유지할 수 있다.

⑥ 데이터 보안성을 최대화할 수 있다.

데이터베이스는 데이터의 중요도와 성격에 따라 보안 유지 기능을 제공하며 사용자 권한에 따라 데이터베이스에 접근할 수 있도록 제어하고 접근 목적에 따라 데이터 작업을 제한하도록 한다.

⑦ 데이터를 표준화하여 관리할 수 있다.

데이터베이스는 구성원들이 서로의 데이터를 이해하기 쉽고 효율적으로 처리하기 위해 데이터를 표준화해 사용하며 이를 유지, 관리할 수 있도록 한다.

14.5. 데이터베이스 활용

파일 처리 시스템의 한계점을 보완하면서 등장한 데이터베이스는 데이터의 중복성을 최소화하고 데이터의 독립성과 일관성, 보안성 최대화 등의 장점으로 기업의 경영정보시스템, 비행기의 예약시스템 같은 기존 응용 분야 외에도 데이터마이닝, 데이터웨어하우스, 빅데이터와 같은 새로운 분야에서도 널리 활용되고 있다.

먼저, 데이터마이닝(data mining)은 대규모로 저장된 데이터 내에 존재하는 패턴이나 규칙, 관계 등을 탐색하고 모형화하여 유용한 지식을 추출하는 과정을 말한다. 즉, 대용량 데이터베이스에서 유용한 정보를 추출하는 일련의 분석 과정이라 할 수 있다. 데이터마이닝은 의사결정에 있어 정보에 대한 의존도가 높을 때 매우 효과적인 의사결정 지원 기법으로 방대한 원천 데이터와 자료를 관리하기 위한 관계형 데이터베이스 관리 시스템과 OLAP가 있어야 한다.

주로 대용량 관측 자료를 이용하며 이론보다는 실무 위주의 컴퓨터 및 데이터 처리 방식이다. 경험적 방법에 근거하는 것이 대부분이며 일반화한 결과를 도출하여 의사결정을 지원하는 것이 주요 목적이다. 데이터마이닝의 주요 기능으로는 분류(classification), 추정(estimation), 예측(prediction), 유사성 집단화(affinity grouping), 연관성 규칙(association rule), 군집화(clustering), 설명(description), 프로파일링(profiling) 등이 있다.

데이터웨어하우스(data warehouse)는 1980년대 중반 IBM이 자사의 제품을 판매하기 위해 처음으로 사용한 개념으로 조직에서 오랜 시간 동안 축적한 내부 데이터와 외부 데이터를 주제별로 통합하여 별도의 조작 없이 여러 각도에서 분석할 수 있게 하는 통합 데이터베이스 시스템을 말한다.

일반적으로 '의사결정 과정을 지원하기 위한 주제 중심적이고 통합적이며 시간성을 가진 비휘발성 자료의 집합'으로 정의한다. 데이터웨어하우스는 의사결정 상황을 지원하기 위해 데이터가 일정 주기로 축적되는 저장소의 역할을 하며 주로 추세분석, 비교분석과 같은 OLAP(On-Line Analytical Processing)를 가능하게 한다.

데이터웨어하우스는 기존의 데이터베이스와 달리 중요한 주제를 중심으로 그 주제와 관련된 데이터로 조직되며 전사적인 관점에서 데이터를 통합하고 특정 시점을 기준으로 시간별로 미리 정의된 데이터만을 저장한다.

사용자의 의사결정을 지원하기 위한 것이 데이터웨어하우스가 존재하는 가장 일차적인 이유라 할 수 있으며 여러 개의 개별적인 운영시스템으로부터 데이터가 집중되어 데이터웨어하우스로 이동되면서 재구조화된다. 마지막으로 빅데이터는 최근 SNS와 같은 다양한 사용자 참여 환경이 사이버 공간에서 제공 되면서 방대한 양의 데이터가 아주 빠르게 확산되고, 쉽게 사라지는 특성을 가지고 있다. 빅데이터는 빅데이터 이해 부분에서 심도 있게 다루고 있다.

15. 빅데이터의 이해

15.1. 빅데이터 개념

가. 빅데이터란?

빅데이터(Bigdata)는 다양한 원천 및 환경에서 수집된 방대한 양의 데이터와 데이터 집합을 의미한다. 즉, 기존에 존재하던 데이터를 처리하는 소프트웨어, 하드웨어 및 분석 프로세스를 사용해서 처리할 수 없을

정도로 방대하고(테라바이트 이상) 복잡한(센서 데이터부터 소셜 미디어 데이터까지) 데이터의 집합을 설명하는 데 사용되는 용어이다.

빅데이터는 비즈니스가 원하는 의미를 갖춘 결과물, 즉 유용한 가치를 창출해야만 한다. 데이터 중심의 비즈니스를 구축하기 위해서는 데이터를 가치 있게 변환하여 실시간 의사 결정을 유도할 수 있는 고급 데이터 분석 기능을 갖추어야 한다. 이는 곧 고급 데이터를 분석하여 의미 있는 결과물을 생산하는 것이 중요해지고 있다는 것을 의미한다.

데이터 그 자체는 가치가 거의 없으나 더 많은 데이터를 수집할수록 조직은 비즈니스와 고객에 대한 통찰력을 얻을 수 있게 된다. 따라서 점점 더 많은 관리자와 팀이 성공적인 의사 결정을 내리고 중요한 직무를 수행하는 데 도움이 되는 빅데이터를 찾고자 한다.

오늘날 계속해서 더 많은 양의 데이터와 더 다양한 유형의 데이터가 매일 수집되고 분석되고 있으며 이를 통해 전 세계의 모든 산업 및 사업 분야의 관리자가 올바른 결정을 내릴 수 있도록 지원하고 있다.

[표 3-2] 빅데이터와 스몰데이터의 차이

| 구분 | 빅데이터 | 스몰데이터 |
|-------------------------|--|---|
| 데이터 원천 (Data Source) | 전통적이지 않으며, 기업 외부에서 생산된 데이터, 소셜 미디어, 센서 데이터, 로그 데이터, 장치 데이터, 비디오, 이미지 등 | 기존의 기업 데이터: 기업 자원 계획 계약 데이터, 고객 관리 시스템, 재무적 데이터 등 |
| 규모 (Volume) | 테라바이트(Terabytes) 페타바이트(Petabytes) 엑사바이트(Exabytes) 제타바이트(Zettabytes) | 기가바이트(Gigabytes) 테라바이트(Terabytes) |
| 속도 (Velocity) | 대부분 실시간 즉각적인 반응 요구 | 자료를 모았다가 일괄 처리 실시간에 근접 즉각적인 반응을 요구하지 않음 |
| 다양성 (Variety) | 정형(structured) 비정형(unstructured) 멀티형(multistructred) | 정형(structured) 비정형(unstructured) |
| 가치 (Value) | 복잡함, 향상된, 비즈니스 분석을 예측, 통찰력 등 | 비즈니스 인텔리전스, 분석, 보고 |

빅데이터는 기존의 스몰데이터(small data)와 여러 측면에서 차이가 있다. 데이터는 기본적으로 다양한 방법으로 수집된 여러 가지 형태의 데이터의 집합을 의미하지만, 빅데이터는 비정형 데이터에 더욱

초점을 맞춘다. 또한, 빅데이터는 기존에 데이터를 분석하던 프로그램 및 시스템으로는 데이터를 분석할 수 없고 병렬 처리를 위한 컴퓨터 시스템들이 필요할 정도로 큰 데이터 집합을 갖고 있다. 빅데이터는 단순히 데이터만 가지고는 의미가 거의 없고 분석을 통한 가치를 창출하는 것에 의의가 있다. 마지막으로 수집된 빅데이터는 단기간 활용하는 것보다 장기적인 활용에 초점을 맞추고 있다.

15.2. 빅데이터의 원천

빅데이터를 수집할 수 있는 원천은 다양하다. 소비자 측면에서 애플 워치와 같은 웨어러블 기술, 스마트폰과 소셜 미디어의 게시물들은 빅데이터의 양(volume)이 증가시키는 중요한 요소이다.



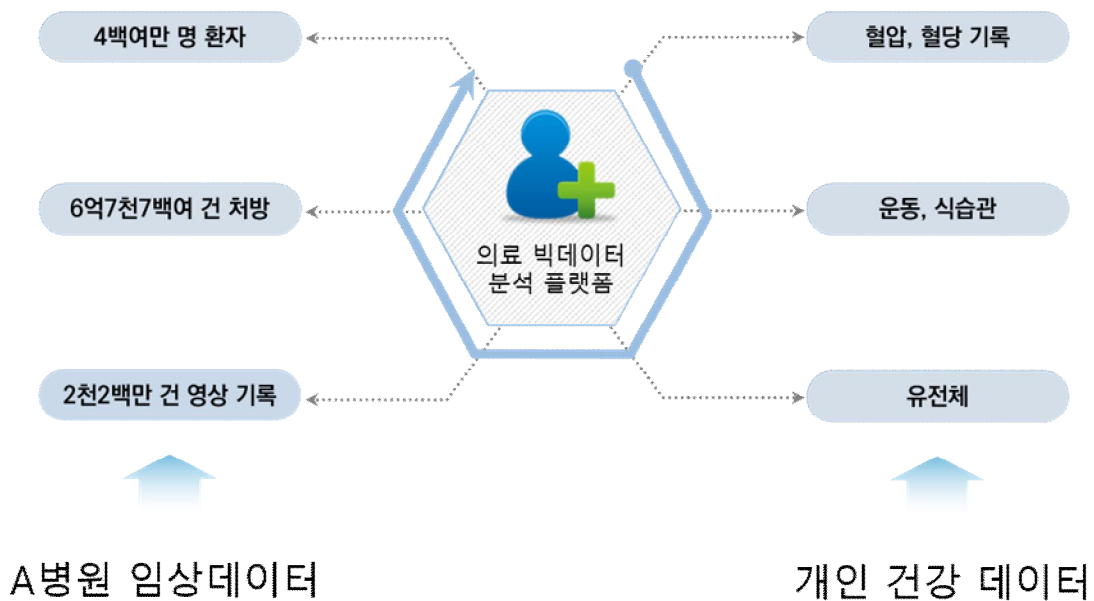
[그림 3-6] 빅데이터 원천

공공 부문과 기업 측면에서는 센서 데이터와 사물인터넷(IoT, Internet of Things)이 자동화된 공장 및 물류 센터, 자율 주행 차량 및 스마트 시티와 같은 IT를 기반으로 한 비즈니스 프로세스를 발전시키는 데 사용되고 있다. 또한, 조직은 생산 프로세스의 일부인 비즈니스 어플리케이션, 소셜 미디어, 센서 및 컨트롤러, 공장 및 사무실의 물리적 환경을 관리하는 시스템, 미디어 자원(오디오, 비디오 방송) 등 다양한 원천에서 데이터를 수집할 수 있다. 의료 부문에서도 의료 기관에서 빅데이터와 빅데이터 분석을 활용하는 노력이 있다. 의료 기관은 수익 주기를 관리, 자원을 활용, 부정행위를 방지, 건강관리 및 서비스 품질 개선을 지원하는 데에 데이터 분석을 이용할 수 있다. 이렇게 수집되는 데이터의 대부분은 비정형(unstructured) 데이터로 이루어져 있으며 기존의 관계형 데이터베이스 관리 시스템으로 처리하기에는 적합하지 않다.

15.3. 빅데이터의 활용

조직은 일상적인 운영, 전략 계획 및 의사 결정을 개선하기 위해 빅데이터를 사용한다. 예를 들어 소매업자는 페이스북, 구글, 링크드인, 트위터, 인스타그램과 같은 소셜 네트워크를 모니터링 함으로써 브랜드에 긍정적인 소비자들의 참여를 높이고, 부정적인 소비자를 식별하여 그들을 대상으로 마케팅을 시도하기도 하며, 최근에는 인플루언서를 통해 그들의 제품을 팔 수 있도록 한다. 광고 및 마케팅 대행사들은 소셜 미디어의 댓글을 추적하여 광고, 캠페인 및 프로모션에 대한 소비자의 반응을 파악하기도 한다.

병원 및 의료 분야에서는 의료 데이터와 환자 기록을 분석하여 퇴원 후 재입원이 필요한 환자들을 식별할 수 있다. B2C(Business-to-customer) 기업은 소셜 네트워크를 모니터링하여 고객의 행동, 제품 인식에 대한 고객 반응을 얻고 제품, 서비스 및 광고에 수정이 필요한 부분을 파악한다. 금융 서비스의 경우 고객 데이터를 활용하여 더욱 정교한 타겟(target) 고객에게 상품을 제안할 수 있다. 제조업체는 예를 들어 장비가 마모되면서 약간씩 달라지는 장비의 미세한 진동 데이터를 분석하여 장비의 유지보수 및 장비를 교체할 최적의 시간을 예측함으로써 비용이 많이 드는 수리 또는 잠재적으로 치명적인 고장을 방지한다.



[그림 3-7] 의료 빅데이터 분석 플랫폼

15.4. 빅데이터 특징

빅데이터라는 개념이 만들어졌을 당시, 기존에 만들어진 데이터와 구분되는 3가지 특징이 있다. 이 세 가지 특징은 Volume(규모), Variety(다양성), Velocity(속도)이다. 최근에는 빅데이터를 통한 가치 창출의 중요성이 더 커지면서 Value(가치), Veracity(진실성), Validity(정확성), Volatility(휘발성) 등의 새로운 V도 등장하였다. 빅데이터가 점차 중요해지면서 이를 설명하는 특징 및 범위가 점차 커지고 있으며 조직은 데이터 분석의 효과를 높이기 위해 빅데이터의 특징들을 잘 파악해야 한다.

가. Volume(규모)

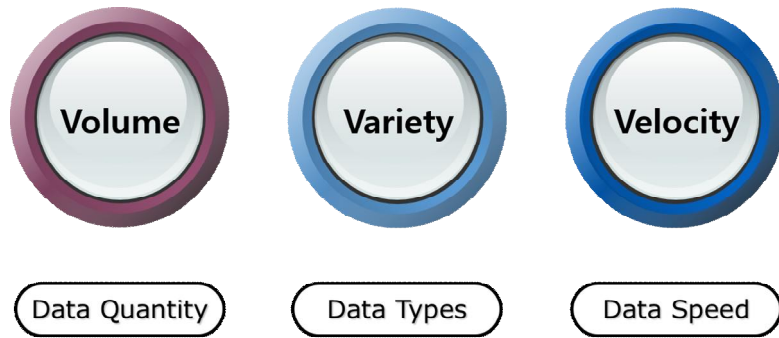
빅데이터는 말그대로 물리적인 크기와 개념의 범위가 대규모이다. 기존의 데이터는 사람 및 조직의 관계에 따라 사람이 데이터를 만들었으나, 현재의 빅데이터는 기계에서 자동으로 생산된다. 예를 들어 소셜 미디어의 게시물 정도는 사람이 일일이 모으는 것이 아니라 여러 플랫폼에 자동으로 저장되고 있다. 빅데이터는 과거의 관계형 데이터, 텍스트 데이터를 포함하여 다양하고 큰 규모의 비정형 데이터(예: 소셜 미디어 게시물, 영상, 사진, 생체 기록 등)를 자동으로 저장하기 때문에 데이터의 크기가 기본적으로 커지는 것이다.

나. Variety(다양성)

오늘날의 데이터는 다양한 형식으로 제공된다. 일부 데이터는 구조화된(structured) 데이터라고 부르는 것으로, 전통적인 데이터베이스에 잘 맞는다. 예를 들어, 기업이 관리하는 고객명, 제품명, 재고, 재무 및 직원 데이터를 포함하는 많은 데이터는 일반적으로 구조화된 데이터라고 한다. 그러나 조직이 처리해야 하는 대부분의 데이터는 기존에 정의된 방식으로 구성되지 않은 비정형(unstructured) 데이터이다. 이러한 비정형 데이터는 정해진 규칙이 없고 연산도 불가능하다. 예를 들어 워드 프로세싱 문서, 이메일, 사진, 비디오 및 전화 메시지와 같은 자원이 비정형 데이터이다. 따라서 분석의 대상이 되는 빅데이터는 한 기업의 데이터 웨어하우스에 있는 정형화된 데이터에서부터 소셜 미디어 게시물, 트윗, 센서 데이터 및 고객 서비스 통화 내역 등과 같은 반정형(semi-structured) 및 비정형(unstructured) 데이터까지 다양하게 확장될 수 있다. 즉 빅데이터는 기존 정형 데이터와 사진, 동영상 등의 비정형 데이터를 모두 포함하고 있다.

다. Velocity(속도)

속도는 데이터를 저장하고 분석하여 결과물을 생산하는 속도를 의미한다. 빅데이터는 실시간으로 생산되며 처리 및 유통 속도까지 매우 빠르다. 데이터가 적시에 처리되지 않으면 해당 데이터는 더이상 정확하거나 유용한 데이터가 아니게 될 가능성이 있다. 예를 들어, 재고의 변동성으로 인해 소비자에게 약속한 대로 주문한 품목을 배송할 수 없을 때 소비자가 실망하지 않도록 재고 관련 데이터를 실시간으로 업데이트해야 한다.



[그림 3-8] 빅데이터의 새로운 특징

라. Veracity(진실성)

데이터의 진실성은 신뢰성과 관련된 특성으로 데이터를 분석해서 나온 일정한 패턴을 신뢰할 수 있도록 해야한다는 의미이다. 데이터의 규모가 크고 다양하다는 것은 반대로 처리하기가 더 어려운 데이터가 포함되었다는 의미이기도 하다. 불완전하거나 누락되거나 중복된 데이터는 데이터 분석으로 추출된 일정한 패턴의 신뢰성을 떨어뜨린다. 만약 잘못된 데이터 세트를 사용한 잘못된 분석을 사용한다면, 분석 결과는 무의미할 것이다. 따라서 관리자는 해당 빅데이터가 진실성을 가지는지 살펴볼 필요가 있으며 동시에 데이터를 해석하는 방법을 이해하기 위한 통찰이 필요하다.

마. Value(가치)

빅데이터를 분석함으로써 비즈니스에 도움이 되는 가치를 창출해낼 수 있어야 한다. 따라서 데이터를 수집하기 전에 해당 데이터를 통해 어떤 의미 있는 생산물을 만들어낼 수 있는지 고민해야 한다.

바. Validity(정확성)

정확성은 신뢰성과 비슷한 개념으로, 데이터가 얼마나 정확한지에 대한 특성이다. Veracity는 데이터에 오류 및 누락된 부분이 없어야 한다는 의미인 반면, Validity는 분석자가 원하는 목표를 해당 빅데이터가 제대로 설명해야 한다는 의미이다. 데이터의 Validity가 없다면 데이터가 아무리 규모가 크고 오류가 없더라도 분석 목표에 맞지 않는 정보이기 때문에 가치가 없어진다.

사. Volatility(휘발성)

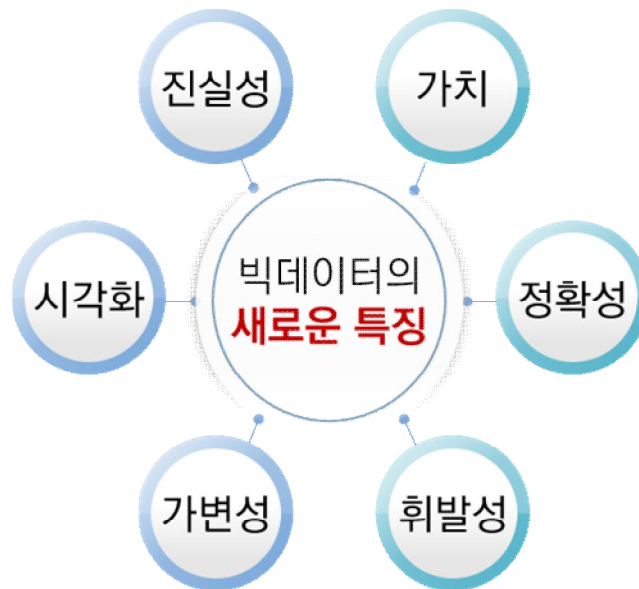
빅데이터는 단기적으로 활용되기 보다 장기적으로 유용한 가치를 만들어낼 수 있어야 한다. 그러기 위해서는 먼저 데이터가 오랫동안 저장될 수 있어야 한다. 즉 시간이 지나면 의미가 없어 장기적으로 활용 가치가 떨어지는 데이터나, 규모에 비해 창출해내는 가치가 떨어지는 데이터는 활용을 고민해보아야 한다.

사. Variability(가변성)

가변성은 데이터의 의미가 변한다는 것을 의미한다. 즉, 빅데이터는 최초 데이터가 생성 되는 시점의 의미가 시간이 지남에 따라 다른 의미로 변한다. 또는 생성된 데이터의 의미가 분석 과정을 거치면서 다른 의미로 해석이 되기도 한다.

아. Visualization(시각화)

빅데이터는 시각화가 가능하다. 데이터를 어떻게 시각화 하나에 따라 데이터 분석 결과의 활용도가 달라 질 수 있다. 즉, 빅데이터는 사용 목적에 따라 다양한 방법으로 시각화가 가능하다. 시각화는 비록 빅데이터 전문가가 아니어도 데이터 분석 결과를 쉽게 이해 할 수 있게 해 주며, 빅데이터 분석 전반적 과정에서 매우 중요한 과정으로 자리 잡고 있다.



[그림 3-9] 빅데이터의 새로운 특징

15.5. 해양물류 빅데이터 응용

가. 해양 물류 빅데이터 활용

해운 경기는 변동성, 불확실성, 복잡성의 확대로 인해 전통적인 방법으로 해운 시황을 분석하기에 한계가 있다. 마찬가지로 기존에는 다양한 해양수산물정보는 검색하기가 어렵고 정보관리가 비효율적이며 정보가 부분적으로 개방, 비합리적인 행동 등으로 활용하기 어려워 다양한 서비스 및 새로운 사업 창출에 한계가 있었다.

현재는 전통적인 예측 방법을 보완할 수 있는 빅데이터, 인공지능 등을 활용한 분석이 활용되고 있다. 운임, 유가, 환율 등을 담은 다양한 시계열 자료와 인공지능의 결합은 시황예측의 정확도를 향상시킬 수 있다. 기존에는 비정형 데이터(기사, SNS), 자동식별시스템(AIS)을 활용하지 않았지만, 이러한 자료를 활용하여 시장 분석도 가능하다.

또한, 실시간 선박 위치정보는 해운시장의 공급 상황을 분석할 수 있도록 하였고, 선박 운항을 통해 얻은 자료(풍향, 조류, 연료소비량 등)를 이용하면 최적의 운항 경로나 속도를 제시하여 운영 효율성을 향상시킬 수 있다. 선박 고장, 선박충돌 사전 정보, 사이버 보안 등 해상물류에서 빅데이터가 활용되는 방안은 다양하다고 할 수 있다.

나. 해양 물류 빅데이터 활용 현황

세계 1위 선사인 머스크(Maersk)는 4차 산업에 대응하기 위해 디지털 본부를 설립하여 운영하면서 빅데이터 분석을 통해 선박의 효율성을 최대 7~8%까지 증가시키는 등 운영의 위험을 최소화하려는 노력을 하고 있다. 해양 물류에서 시황예측이 중요한 만큼 일본해운 3사는 기계 학습을 기반으로 시황예측 모델을 개발하였고 상하이국제항운중심(Shanghai International Shipping Institute, SISI)은 대학과 협업하여 인공지능 기반 해운시황예측모델을 개발 중이다.

국내에서는 해양수산부는 해양수산분야의 다양한 기관에 산재된 빅데이터를 유기적으로 연결하는 데이터 허브의 역할을 위해 산하의 해양수산 빅데이터 플랫폼을 구축하였다. 해당 플랫폼은 다양한 기관의 연계와 통합을 통해 해양수산정보의 가치를 극대화하였다.

통합 포털에서 데이터 맵을 기반으로 해양수산 빅데이터를 검색할 수 있으며 웹 데이터 등의 비정형 데이터를 실시간으로 분석하고 해양 공간데이터 시각화, 통계, 연산, 분석을 쉽고 편리하게 할 수 있다. 예를 들어 센터 참여 기관들은 고품질의 항만 데이터, 해양 환경, 해양관측 정보, 선박기관 고장 데이터, 선박검사, 일일운항실적, 운항통제 등의 데이터들을 보유하고 있다.

다. 해양 물류 빅데이터 활용

해양 물류에서 기존에 존재하는 데이터와 IoT 및 모바일을 기반으로 실시간으로 데이터를 융합하여 실시간 항만물류 정보 서비스, 생산 및 시스템 구축, 해상 교통안전 강화, 선박운항 안전 강화, 선박 검사, 일일운항실적, 운항통제데이터 서비스 등을 제공하고 있다.

특히, 빅데이터를 활용하여 다양한 해운 시황을 예측할 수 있게 되었다. 이를 통해 항만물류의 효율과 생산성 및 안정성을 높이고 물류 경쟁력을 강화하고 있다.

해양 물류와 관련된 빅데이터의 종류에는 선박 운항데이터, 글로벌 해상 기상데이터, 조선기자재 데이터, 어업별 어선 위치정보 데이터, 해양사고 정보 통계데이터, 선박 검사 데이터, 해양 기상 현황 데이터, 터미널에 수출 및 수입된 컨테이너 데이터, 운송 차량에 의한 컨테이너 운송 데이터, 컨테이너 취급 장비의 터미널 이동 및 야적 관련 데이터, 운송 차량의 위치 정보 데이터, 크레인 및 터미널 전용 차량 등 하역장비의 실시간 위치 및 시그널 데이터 등이 있다.

▶ 선박 공급 예측

자동식별시스템(Auto Identification System, AIS) 자료를 이용하여 특정 지리적 공간에 위치하거나 공간을 통과한 선박의 위치를 실시간으로 분석함으로써 시장에 공급되는 선박을 사전에 예측할 수 있다.

▶ 시황 예측

텍스트 마이닝(감성분석 등), 관련 산업의 기사, SNS 자료 등을 활용하여 시장참여자들의 감정 및 의견을 파악할 수 있다.

▶ 최적 항로 분석

선박 운항을 통해 얻은 자료(풍향, 조류, 연료소비량 등)를 이용하면 최적의 운항 경로나 속도를 제시할 수 있다.



[그림 3-10] 지능형 선박관리 시스템-최적항로 시스템

▶ 선박사고 예방

선박 엔진이나 운항데이터의 고장과 관련된 데이터 및 선박 검사 데이터 등을 활용하여 운항 전 선박 사고를 예측하고 예방할 수 있다.

▶ 선박 운항 안정성 검증

선박을 운항하는 해운사나 선박관리사의 선박 운항데이터와 전자해도 데이터를 활용하여 선박 운항에 대한 안정성을 검증할 수 있다.

▶ 해양사고 저감 서비스

해양사고 정보 및 선박 검사 데이터, 기상 현황 데이터, AIS 데이터 등을 활용하여 선박 이동의 예상 경로에 따른 사고를 예방할 수 있다.



[그림 3-11] 선박의 운항데이터와 전자해도를 기반으로 한 운항안정성 검증
(출처 : 해양수산개발원)

▶ 선박 연비 분석

선박의 운항데이터와 기상데이터를 활용하여 조선 기자재를 설치한 선박의 연비개선율을 분석할 수 있다.

▶ 탱커선 운항률, 평균 속도, 연료 소비량 분석

선박별, 선대별, 조선소별 해상환경 데이터와 연계된 자료를 분석할 수 있다.

▶ 컨테이너 터미널 혼잡도 분석

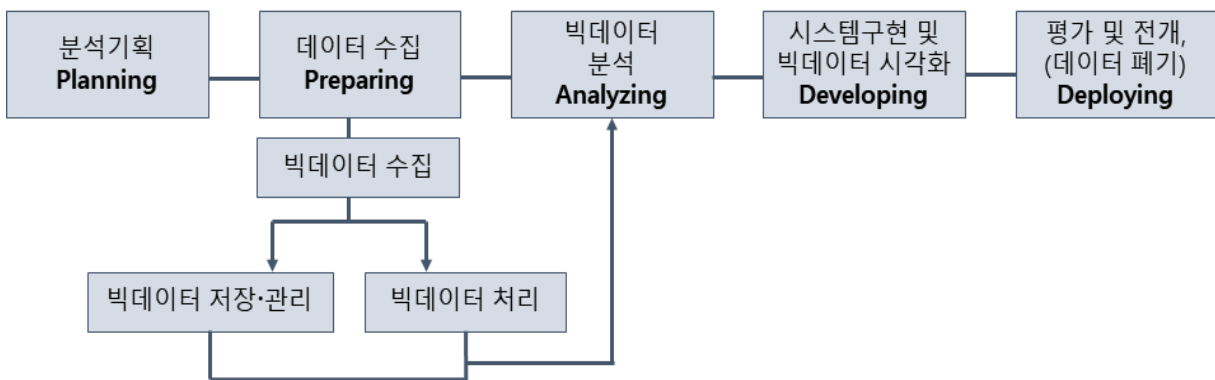
컨테이너 터미널의 CHE 실시간 위치정보 데이터 및 외부차량의 실시간 위치정보 데이터를 통해 터미널 혼잡도를 고려하여 외부차량을 통제하는 시스템을 개발할 수 있으며 VBS(Vehicle Booking System)과 연계할 수 있다.

▶ 전국 항만/터미널의 턴어라운드타임 현황 파악

운송차량 및 컨테이너 터미널의 반출입 차량에 대한 데이터를 통해 작업 시간을 파악할 수 있고 요일별, 선박별, 유형별(수출/수입/수추립/환적) 현황을 빠르게 파악할 수 있다.

16. 빅데이터 분석 프로세스

빅데이터 분석의 주요 관점은 새로운 종류의 데이터나 많은 양의 트랜잭션 데이터(transaction data), 다양한 유형의 데이터 내에 숨겨진 패턴, 상관관계를 발견해 의사 결정에 활용하기 위함이다. 이러한 과정에는 데이터 수집부터 최종 시각화를 통해 데이터 폐기 까지 일련의 과정이 필요하다. 이러한 과정을 통해 창출된 유용하고 가치 있는 정보 즉, 인사이트(Insight)는 조직의 의사결정에 매우 중요한 역할을 한다. 따라서 효과적인 데이터 분석을 하기 위해서는 체계적인 절차와 방법이 정리된 데이터 분석이 중요하다. 빅데이터 분석 프로세스에는 여러 단계를 거쳐야 한다.



[그림 3-12] 빅데이터 분석 프로세스

16.1. 빅데이터 수집

빅데이터 수집은 단순한 개념으로 다양한 데이터를 분석 목적으로 수집한다 할 수 있다. 하지만 좀 더 광의적 개념의 빅데이터 수집은 단순히 데이터를 확보하는 것이 아니라, 필요한 데이터를 다양한 소스로 부터 검색해 수집한 후 필요한 형태로 변환하는 과정을 통해 정제된 데이터를 확보하는 과정을 의미하기도 한다. 빅데이터 수집은 보통 분석의 목적과 대상을 우선적으로 설정한 후 이 목적을 달성하기 위해 필요한 데이터를 선정해 수집하게 된다. 즉, 분석에 대한 기획을 한 후 이에 적합한 데이터를 준비해야 하며, 이러한 준비를 통해 데이터를 수집해야 한다. 특히, 데이터 선정에 있어서 다음의 몇 가지 고려사항이 있다.

① 수집가능성

수집가능성은 데이터 선정 시 가장 우선하여 고려해야 할 사항이다. 수집의 가능성 여부를 파악하고, 비록 수집할 수 있더라도 데이터 활용 측면에서 향후 많은 비용이 필요하다면 이 데이터는 수집가능성이 높지 않다고 할 수 있다.

② 보안문제

수집한 데이터가 개인정보 또는 저작권 문제의 유무에 대해 살펴보아야 한다. 공식적인 동의 없이 개인정보 데이터를 수집해 분석하면 향후 법적으로 문제가 될 수 있다.

③ 데이터 정확성

수집한 데이터의 정확성은 활용목적에 세부항목이 정확히 존재하는지에 대하여 검토해야 한다. 수집 목적에 맞는 데이터를 확보하기 위해서는 사전처리 과정과 데이터의 사후처리 방안도 마련되어야 한다.

④ 데이터 난이도

데이터 수집 및 처리에 들어가는 구축비용이 많이 들어갈 경우와 데이터 수집의 분석설계가 필요한 데이터를 얻기 위해 많은 정제 과정이 필요한 경우로 구분하여 대안을 마련해야 한다.

⑤ 수집비용

데이터를 수집하기 위해 직접적으로 들어가는 비용으로 아무리 분석에 필요한 데이터라도 수집비용이 예산 범위 밖에 있다면, 수집이 쉽지 않을 수 있다.



[그림 3-13] 빅데이터 분석기획 및 데이터준비 세부단계

가. 데이터 수집 프로세스

의사결정에 필요한 데이터 수집은 빅데이터 분석에 있어 매우 중요한 단계이다. 수집된 데이터가 의사 결정과 관계성이 낮거나 잘못된 데이터를 수집해 분석한다면, 이 모든 과정이 헛수고가 될 수 있다. 데이터 분석에 있어 "데이터는 거짓말을 하지 않는다"라는 속설이 있다. 즉, 데이터 수집이 제대로 되었다면, 분석한 결과의 신뢰성에는 문제가 없다는 것을 의미한다. 이러한 데이터 수집절차를 설계하기 위해서 수집데이터 도출, 목록 작성, 데이터 소유기관 파악 및 협의, 데이터 유형 분류 및 확인, 수집기술 선정, 수집계획서 작성, 수집 주기 결정, 데이터 수집실행 과정이 있다.

① 수집데이터 도출

데이터 분석 목적에 맞는 데이터 도출이 필요하다.

② 목록작성

수집가능성, 세부 데이터 항목, 보안 문제, 비용을 고려한 데이터 수집 목록을 작성한다.

③ 데이터 소유기관 파악 및 협의

데이터 소유자의 데이터 개발 현황, 적용기술, 보안 등을 파악 및 데이터 수집 관련 보안 사항, 개인정보보호 관련 문제 등 점검한다.

④ 데이터 유형 분류 및 확인

수집 대상 데이터 유형을 분류하고 데이터 포맷 등을 확인한다.

⑤ 수집 기술 선정

데이터 유형 및 포맷 등에 맞는 수집기술을 설정하고, 수집기술은 데이터 소스로부터 다양한 유형의 데이터를 수집하기 위해 확장성, 안정성, 실시간성 및 유연성 확보가 필요하다.

⑥ 수집계획서 작성

수집 대상 데이터 출처, 수집기술, 수집 주기 등을 반영한 계획서를 작성한다.

⑦ 수집 주기 결정

데이터 유형에 따라 배치 또는 실시간 방식을 적용한다.

㉔ 데이터 수집실행

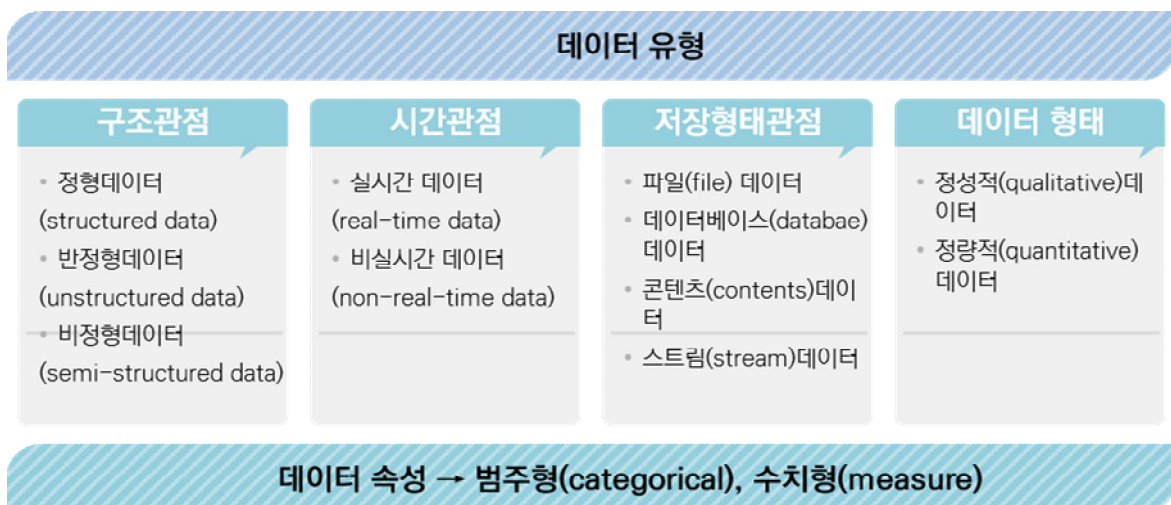
사전 테스트를 진행하고 데이터 수집을 시행한다.

나. 수집데이터의 대상

데이터를 선정한 후 선정된 데이터 위치를 파악하고, 데이터 유형을 이해해 수집 시 적용할 기술 및 보안 사항 등을 점검하고, 수집계획서를 작성한다. 수집데이터 대상은 크게 내부데이터와 외부 데이터로 나눌 수 있다. 내부데이터인 경우, 내부 시스템과 데이터 연계가능여부, 데이터 종류, 데이터 수집주기 등을 파악한다. 이러한 내부데이터에는 서비스(SCM, ERP, CRM, 거래시스템), 네트워크(방화벽, 백본, IPS), 그리고 마케팅(VOC접수) 등이 있다. 반면, 외부데이터인 경우, Open API는 개방하는 데이터의 종류 및 형태를 파악해 데이터의 양과 트래픽의 정도를 확인한다. 외부데이터로는 소셜네트워크 데이터, 네트워크 데이터(장비간발생로그, M2M, 센서데이터 등), 그리고 공공데이터(정부 공개 공공데이터) 등이 있다.

다. 데이터 유형 및 속성

데이터 유형에는 정형데이터, 반정형 데이터, 비정형 데이터를 포함한 구조관점의 데이터 유형이 있고, 실시간 데이터와 비실시간 데이터를 포함한 시간 관점의 데이터 유형이 있다. 또한, 저장형태에 따라서 파일 데이터, 데이터베이스 데이터, 콘텐츠 데이터, 스트림 데이터로 데이터가 저장되는 방식에 따라 다양하게 분류가 된다. 그 외 데이터는 언어, 문자 등의 형태인 정성적 데이터와 도형, 수치, 기호 등과 같은 정량적 데이터가 있다. 이러한 데이터의 특성을 알려 주는 속성은 범주형과 수치형이 있다. 범주형 데이터는 조사 대상을 특징에 따라 범주로 구분하여 측정되는 변수이며, 명목형과 순서형인 데이터 종류를 포함하고 있다. 수치형 데이터는 숫자를 이용해 표현할 수 있는 데이터이고 이산형과 연속형인 데이터 종류를 포함하고 있다.



[그림 3-14] 빅데이터 유형 및 속성

라. 데이터 수집 방식 및 기술

다양한 데이터 유형 중 데이터 수집 방식과 이와 관련된 기술은 데이터의 구조적 관점에서 살펴볼 수 있다. 즉, 정형, 비정형 그리고 반정형 데이터의 종류에 따라 수집하는 기술의 다양성이 존재한다.

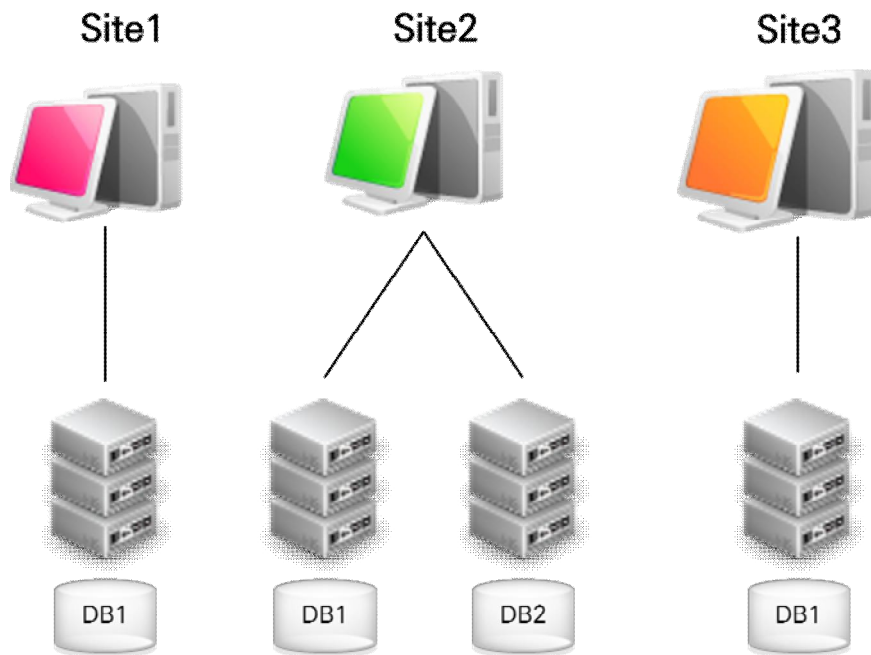
예를 들면, 정형데이터는 일반적으로 많은 조직에서 사용하고 있는 관계형 데이터베이스나, 엑셀 스프레드 시트에 저장된 데이터로 수집대상 데이터를 추출, 가공하여 데이터 웨어하우스(DW) 및 데이터 마트(DM)에 저장하는 기술인 ETL(extraction, transformation, load의 약자)이나 TCP/IP 기반으로 파일을 송·수신하는 응용계층 통신 프로그램인 FTP(file transfer protocol)와 같은 기술로 수집할 수 있다. 반면, 웹문서, SNS 데이터 등의 반 혹은 비정형데이터는 웹크로링 기술을 사용해 데이터를 수집할 수 있다.

[표 3-3] 구조적 관점의 빅데이터 유형별 종류와 수집기술

| 데이터 유형 | 개념 | 데이터 종류 | 수집기술 |
|--------|---|-------------------------------------|---|
| 정형데이터 | 미리 정해놓은 형식과 구조에 따라 저장되도록 구성하여 필드에 저장된 데이터 | RDB, 엑셀 스프레드 시트, 구글 스프레드시트 | ETL, FTP, Oen API |
| 반정형데이터 | 데이터의 형식과 구조가 변경될 수 있는 데이터 | HTML, XML, JSON, 웹문서, 웹로그, 센서 데이터 등 | Crawling, RSS, Open API, FTP |
| 비정형데이터 | 정형화되지 않은 데이터, 정형데이터와 반대로 형태가 없고 연산할 수 없는 데이터. | 소셜 데이터, 문서(워드, 한글), 이미지, 오디오, 비디오 등 | Crawling, RSS, Open API, Streaming, FTP |

16.2. 빅데이터 저장 및 관리

데이터 수집을 통해 습득된 빅데이터를 유용한 정보로 추출하려면, 빅데이터를 효과적으로 저장(storage)과 보관(archive)해야 한다. 빅데이터 저장은 수집한 데이터를 분산 스토리지(distributed storage)에 적합한 방법으로 안전하게 그리고 보전 가능한 방법으로 보관하는 것이다. 따라서 데이터 백업, 입출력, 보안 관리 기능 등을 가진 분산파일 저장시스템이나 분산데이터베이스 관리기술이 반드시 요구된다. 클라우드 스토리지를 이용한 빅데이터 저장이 일반적 전략인데 이는 클라우드 기술이 가지고 있는 대규모성, 신뢰성, 자동고장 복구 등의 특징이기 때문이다.



[그림 3-15] 분산 데이터 관리 시스템

또한, 빅데이터 저장 및 관리는 대용량 데이터를 저장하고 관리하는 시스템이기 때문에 대용량의 저장공간, 확장성, 가용성, 신뢰성 등을 보장해야 한다. 하지만, 빅데이터 분산 저장소가 아닌 일반적인 데이터웨어하우스는 실시간 증권 데이터 활동의 경우 등과 같은 최신정보를 자주 업데이트 해야 하는 환경에서는 대용량 데이터를 처리하지 못할 수 있다. 또한, 비정형, 반정형 데이터 형식은 구조화된 관계형 데이터베이스를 바탕으로 한 기존 데이터웨어하우스 저장은 적절하지 않다.

가. 데이터 저장 방법

보통 데이터를 저장하기 위해서는 컴퓨터 하드디스크 드라이브(HHD)나 Solid State Drive(SSD) 또는 USB나 외장하드드라이브와 같은 외장 장치 또는 클라우드 등을 이용한다. 하지만, 빅데이터는 대용량 데이터를 조직단위로 저장하고, 관리하기 위해서 효율성이 매우 중요하다. 이에 정형데이터와 같은 구조가 있는 데이터는 보통 데이터베이스에 데이터를 일정한 규격에 맞게 처리하고 DBMS(Data Base Management System)를 통해 효율적으로 관리한다. DBMS란 데이터베이스를 관리하기 위한 툴, 검색과 저장하는 기능을 제공한다.

[표 3-4] DBMS 종류

| 제품 | 특징 |
|--------|---|
| Oracle | <ul style="list-style-type: none"> • 오라클에서 만들어 판매 중인 상업용 데이터베이스 • MySQL, MSSQL보다 대량의 데이터 처리에 용이 • 비공개 소스에 폐쇄적인 운영 및 고비용 |
| MySQL | <ul style="list-style-type: none"> • MySQL사에서 개발하고 현재 오라클에 합병됨 • 오픈소스로 제공되는 무료 프로그램 • 5천만건 이하의 데이터를 다루기에 적합함 |
| MSSQL | <ul style="list-style-type: none"> • 마이크로소프트사에서 개발한 상업용 데이터베이스 • 비공개 소스에 폐쇄적인 운영 |
| Tibero | <ul style="list-style-type: none"> • 티맥스 데이터에서 제작한 국내 DBMS • SQL 등을 포함한 오라클의 제품과 거의 동일한 호환성을 제공함 • 자기튜닝을 통한 성능 최적화, 지속적인 DB모니터링, 성능 관리 지원 등을 제공함 |

또한, DBMS는 여러 사람이 공유하고 사용할 목적으로 관리되는 데이터베이스를 조작·접근·복구기능·보안 기능을 하는 소프트웨어라고 이해할 수 있다. DBMS의 장점은 데이터 자료의 통합성과 데이터 접근성이 쉬워진다. 또 중복데이터를 제거하고 쉽게 정리해서 유지보수가 쉽고 표준화가 가능하다. 따라서 응용 프로그램의 개발비용이 감소한다. 하지만, DBMS는 운영비용이 많이 들고 자료 백업 및 회복방법이 복잡하다는 단점이 있다.

나. 빅데이터 저장기술

기존의 정형 데이터를 저장하는 기술과 다르게 빅데이터는 데이터 용량이 크다는 이유로 기존의 데이터 저장 파일 시스템에 저장하면 관리의 효율성과 성능이 저하된다. 따라서 저장기술 및 관리가 다르게 요구된다. 빅데이터와 같은 대용량 데이터 저장기술에는 분산파일 시스템, 데이터베이스 클러스터, NoSQL, 병렬 DBMS, 네트워크 구성 저장시스템, 클라우드 파일 저장시스템이 있다.

분산파일 시스템은 컴퓨터 네트워크를 통해서 여러 호스트 컴퓨터와 공유하는 파일에 접근할 수 있게 하는 호스트 서버 기반의 파일 시스템이다. 데이터베이스 클러스터는 관계형 데이터베이스 관리 시스템으로써 여러 개의 서버상에 하나의 데이터베이스를 구축하는 시스템이다.

NoSQL은 일반적인 RDBMS와 다른 DBMS를 지칭하기 위한 용어로 데이터 저장에 고정된 테이블 스키마가 필요하지 않고 조인 연산을 사용할 수 없고, 수평적으로 확장이 가능한 DBMS 시스템이다. 병렬 DBMS는 다수의 마이크로프로세서를 사용하여 여러 디스크에 대한 입출력, 갱신 등의 데이터베이스를 동시에 처리하는 데이터베이스 시스템이다.

네트워크 구성 저장시스템은 상호 다른 데이터 저장 장치를 하나의 데이터 서버에 연결하여 전체적으로 통합한 데이터를 저장 및 관리하는 시스템이다. 마지막으로 클라우드 파일 저장시스템은 클라우드 컴퓨터 환경에서 가상화한 기술을 사용한 분산파일 시스템이다.

또한, 데이터를 저장하기 위해서는 데이터의 용량을 산정해야 한다. 산정을 위해서 완전성, 유일성, 정확성 기준에 따른 데이터 품질을 평가한다.

| [표 3-5] 빅데이터 저장기술 | |
|-------------------|---|
| 기술 | 특징 |
| 분산파일 시스템 | 컴퓨터 네트워크를 통해서 여러 호스트컴퓨터와 공유하는 파일에 접근할 수 있게 하는 호스트 서버 기반의 파일 시스템 |
| 데이터베이스 클러스터 | 관계형 데이터베이스 관리 시스템으로써 여러 개의 서버상에 하나의 데이터베이스를 구축하는 시스템 |
| NoSQL | 전통적인 RDBMS와 다른 DBMS를 지칭하기 위한 용어로 데이터 저장에 고정된 테이블 스키마가 필요하지 않고 조인 연산을 사용할 수 없고, 수평적으로 확장이 가능한 DBMS 시스템 |
| 병렬 DBMS | 다수의 마이크로프로세서를 사용하여 여러 디스크에 대한 입출력, 갱신 등의 데이터 베이스를 동시에 처리하는 데이터베이스 시스템 |
| 네트워크 구성 저장시스템 | 상호 다른 데이터 저장 장치를 하나의 데이터 서버에 연결하여 전체적으로 통합한 데이터를 저장 및 관리하는 시스템 |
| 클라우드 파일 저장시스템 | 클라우드 컴퓨터 환경에서 가상화한 기술을 사용한 분산파일 시스템 |

그 외 대표적인 빅데이터 저장과 관리기술로 하둡(Hadoop)의 분산파일 시스템인 HDFS(Hadoop Distributed File System), Cassandra, MongoDB 등이 있으며, 한국전자통신연구원(ETRI)에서 개발한 GLORY-FS도 있다. 하둡은 대용량 데이터를 분산 처리할 수 있는 자바 기반의 공개 소스 프레임워크다. HDFS에 데이터를 저장하고, 분산 처리 시스템인 맵리듀스(MapReduce)를 이용해 데이터를 처리한다. 대규모 데이터 저장과 처리에 고비용이 요구 된다는 점을 감안 했을 때 오픈 소스인 하둡은 기존 데이터베이스 시스템보다 비용이 적게 들고, 데이터를 분산해 저장해두기 때문에 처리 속도가 빠르다는 장점이 있다.

가. 빅데이터 처리 특징

빅데이터 처리기술은 지속적으로 발생하는 스트림 데이터 및 대규모 저장 데이터의 적시 분석 처리를 지원하기 분산 병렬 데이터 처리(distributed data parallel) 기술이다. 이러한 빅데이터 처리에는 몇 가지 주요 특징을 고려해야 한다.

먼저, 빅데이터 처리는 기존의 데이터 처리 방식과는 다르게 의사 결정에 있어 즉시성이 덜 요구된다. 둘째, 대용량의 데이터를 기반을 둔 분석이 주요 이슈이기 때문에 빅데이터 분석은 장기적 관점에서 전략적이고, 종종 일회성 거래 처리나 행동 분석을 지원해야 한다. 셋째, 단순한 프로세싱 모델이 아닌 다양한 데이터 소스, 복잡한 로직 처리, 대용량 데이터 처리 등을 위해 처리의 복잡도가 가장 높고, 통상적으로 분산 처리 기술이 필요한지 판단해야 한다. 넷째, 빅데이터는 처리해야 할 데이터량이 방대한 것이 일반적일 뿐 아니라 처리 과정 역시 복잡하므로 실시간 또는 준 실시간 처리가 보장되어야 하는 분석에서는 적합하지 않을 수 있다. 마지막으로, 빅데이터는 웹 2.0의 다양한 소셜 미디어 데이터, 로그 파일 스트림 데이터 등의 비정형데이터 비중이 높다는 것에 명심해야 한다.

나. 빅데이터 처리 방식 및 기술

빅데이터를 처리 하는 방법과 관련된 기술은 다양하다. 이러한 처리 방식은 일괄 처리, 실시간 처리 그리고 대화형 처리가 있다. 우선, 빅데이터 일괄 처리(batch processing)기술은 빅데이터를 여러 서버로 분산하여 각 서버에서 나누어 처리하고, 이를 다시 모아서 결과를 정리하는 분산 처리 기술 방식이다.

즉, 초,분,시간,일,월과 같은 특정 시간 단위 동안 데이터를 모아두었다가 한 시점에 순서적으로 처리하는 방식이다. 이러한 방식은 일정한 기간의 데이터를 한번에 처리한다는 장점이 있는 반면, 실시간 데이터 처리 방식이 아니라 즉시에 필요한 실시간 데이터 조회가 힘들다는 특징이 있다. 일괄 처리와 관련된 기술에는 구글 맵리듀스, 하둡 맵리듀스, 마이크로소프트 드라이애드 등이 있다.

다음으로 빅데이터 실시간 처리기술(real-time processing)은 끊임없이 발생하는 데이터 스트림에 대해 실시간으로 처리 하고 분석하는 기술이다. 데이터의 특성에 따라 정형데이터와 비정형데이터 스트림 처리 기술로 나누어 볼 수 있다.

정형 이벤트 스트림 처리 연산을 제공하는 CEP(Complex Event Processing) 기술이 주로 활용되고 있고, 비정형 스트림 및 대규모 스트림 처리를 위해 데이터 처리 로직을 자유롭게 정의하여 사용 가능한 분산 스트림 병렬 처리 기술이 나오기 시작했다.

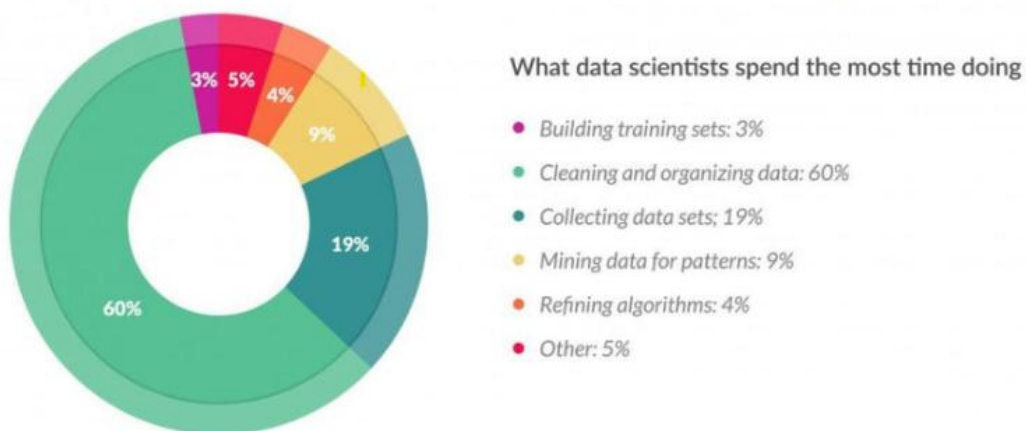
셋째, 대화형 처리(interactive processing)는 REPL(Read-Evaluate-Print Loop)로 하나의 입력(read)을 받아서 처리(evaluate)하고 결과를 반환(print)하는 환경으로 구현된 프로그램을 의미한다. 일반적으로 JavaScript, Python, Scala 등이 대화형 환경을 제공해 준다. 마지막으로, 빅데이터 처리 프로그래밍 지원 기술은 분산 데이터를 처리하는 프로그래밍 언어인 구글의 소재 Sawzall과 병렬처리를 하는 고성능 데이터-플로우 언어와 실행 프레임워크인 하둡 Pig가 있다.

다. 빅데이터 전·후처리 및 저장

빅데이터 저장은 다시 전처리, 후처리 그리고 빅데이터 저장으로 분류 된다. 먼저, 빅데이터 전처리(pre-processing)는 수집한 데이터를 저장소에 적재하기 위해 데이터 필터링, 유형 변환, 정제 등의 기술을 사용하여 데이터를 변환하는 과정이다.

필터링 과정에서는 분석 목적에 부합하지 않은 데이터를 필터링으로 제거해 분석 시간을 줄이고, 저장 공간을 효율적으로 활용 할 수 있게 해준다. 유형 변환은 데이터를 분석에 용이한 형태로 일관성 있는 형태로 변환하는 과정이다.

다음으로 정제는 수집된 데이터에 결측값(missing value)과 노이즈(noise)를 제거하는 과정이다. 일반적으로 빅데이터 전처리 과정은 많은 시간과 노력이 요구된다. 2016년 포버스(Forbes)의 조사에 따르면, 데이터과학자들이 가장 많은 시간을 투자 하는 분야가 데이터 전처리 과정이라고 나타났다. 이와 동시에 데이터과학자들이 가장 싫어하는 과정 역시 데이터 전처리 과정이라고 한다.



[그림 3-16] 데이터과학자들 빅데이터 분석 작업비중

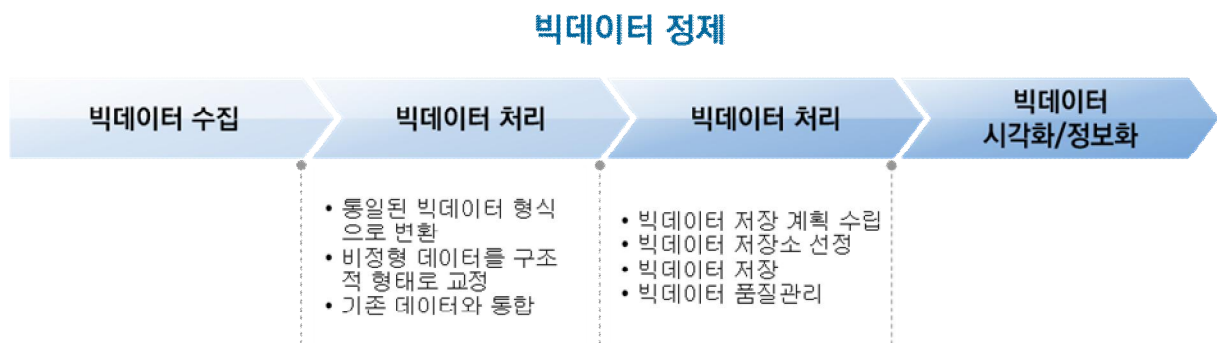
(출처: Press, G.(2016), Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says, Forbes)

빅데이터 후처리(post-processing)는 저장 데이터를 분석에 용이하도록 데이터 변환, 통합, 축소 등을 통해서 분석 효율성을 높이는 과정이다. 후처리에는 데이터 변환, 데이터 통합 그리고 데이터 축소가 있다.

먼저, 데이터 변환은 다양한 형식으로 저장된 데이터를 일관성 있는 형식으로 변환하여 분석에 용이하도록 하는 과정이며, 평활화(smoothing), 집계(aggregation), 일반화(generalization), 정규화(normalization), 속성생성(attribute/feature construction)등을 거치게 된다.

다음으로 데이터 통합은 비록 출처는 다르지만 연관성 있는 데이터들을 하나로 결합하는 기술이다. 마지막으로 데이터 축소는 분석에 불필요한 데이터를 축소함으로써 고유 데이터의 특성을 손상하지 않기 위한 과정이다.

(3)빅데이터 저장은 저장할 데이터의 포맷 등 유형을 검토한 뒤, 데이터에게 유리한 방식으로 저장하는 과정이다.



[그림 3-17] 빅데이터 정제 과정

가. 빅데이터 분석이란

데이터 분석은 데이터 수집 및 처리가 완료된 방대하고 다양한 데이터를 고급 분석 기법을 활용하여 결과를 도출해내는 과정을 의미한다. 고급 분석 기술의 대표적인 예로는 텍스트 분석, 데이터 마이닝, 머신 러닝, 예측 분석, 통계분석 및 자연어 처리 등이 있으며, 기업들은 이러한 고급 분석 기술을 사용하여 데이터 원천에서 독립적인 결과 또는 기존의 데이터와 결합하여 새로운 통찰을 얻게 된다.

나. 데이터 분석 단계

미국의 정보기술 자문회사인 가트너(Gartner)에서는 데이터 분석을 4단계로 구성하여 '분석 가치 에스컬레이터'를 제공하였다. 데이터 분석의 네 단계는 기술(descriptive) 분석, 진단(diagnostic) 분석, 예측(predictive) 분석, 처방(prescriptive) 분석으로 이루어져 있으며 단계가 올라갈수록 분석의 난이도와 얻을 수 있는 가치가 높아짐을 의미한다.

① 기술 분석

가장 기본적인 분석 단계이며, 데이터를 바탕으로 과거에 일어난 일 및 현재 일어나는 일이 무엇인지 살펴볼 수 있다. 이를 통해 기업은 현재의 상황과 문제점을 파악할 수 있다.

② 진단 분석

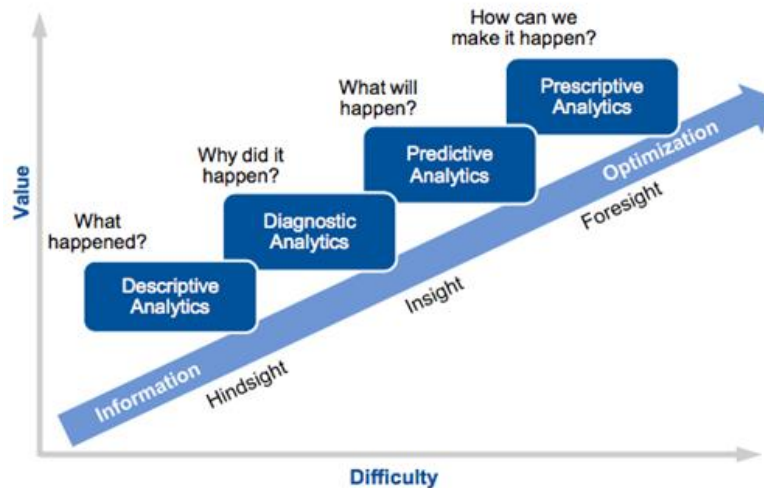
데이터를 통해 알아낸 문제점이나 상황을 진단하는 데 초점이 맞추어진 분석으로, 기술 분석을 통해 알아낸 결과의 원인을 파악하는 분석 단계이다.

③ 예측 분석

과거의 데이터를 바탕으로 일정한 패턴을 분석하여 미래 조직 및 고객의 행동을 예측하는 분석이다.

④ 처방 분석

예측 분석을 바탕으로 최적의 결과를 성취할 최선의 방법을 계산하는 분석이다. 한정된 자원 안에서 최적의 생산물을 찾아 비용을 최소화하고 효율을 최대화하는 해결책을 제시한다.



[그림 3-18] 분석 성숙도 모델(Analytics Ascendancy Model)
(출처 : Gartner, 2012)

다. 빅데이터 분석의 종류

빅데이터의 분석을 통해 현상에 대해 인과 관계 도출, 일정한 패턴 발견, 미래 예측 등이 가능하다. 대표적으로 변수들 사이의 인과 관계를 도출하는 것은 통계 분석, 미래를 예측하는 방법은 기계 학습(머신 러닝), 데이터 마이닝 등을 사용한다. 예를 들어 데이터 마이닝은 데이터 내에 존재하는 일정한 패턴을 발견하는데 사용되는 분석 방법 중의 하나인데 데이터 마이닝 알고리즘을 통해 변수들을 분류할 수 있으며, 계측 및 비계측적으로 객체를 군집화는 물론 알고리즘을 통해 행동 예측이 가능하다. 다음은 빅데이터 분석의 종류에서 설명하고 있다.

① 데이터 마이닝(Data Mining)

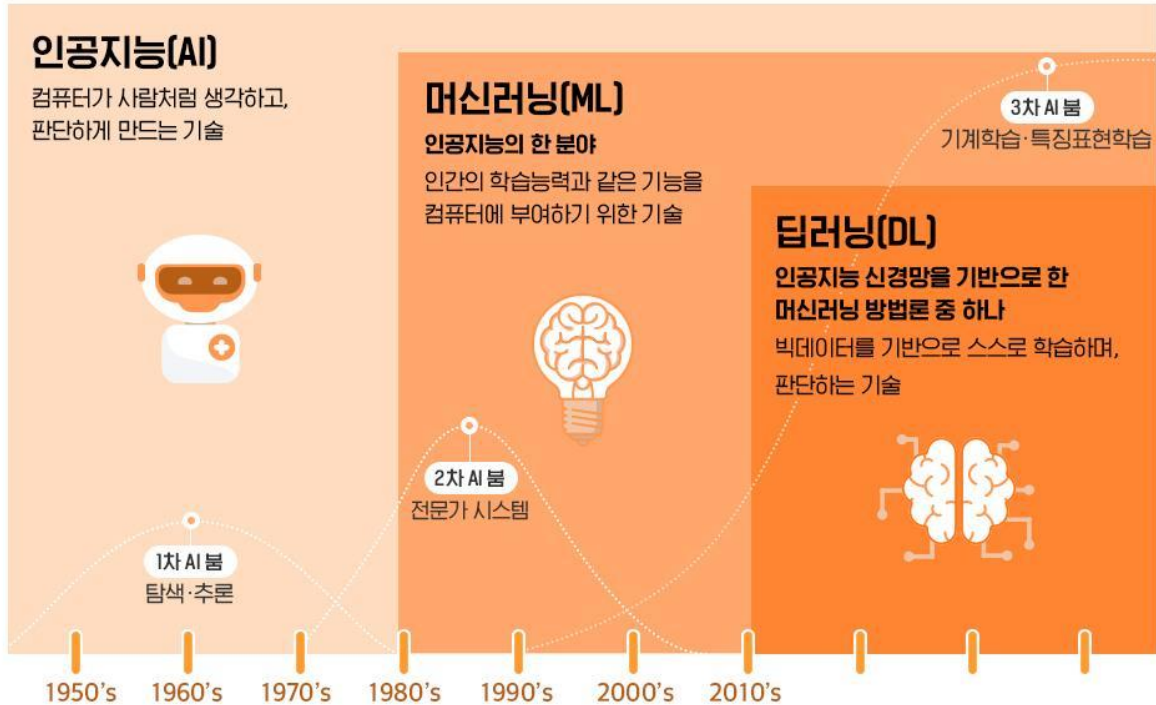
데이터 마이닝은 대규모의 데이터 세트를 바탕으로 데이터 사이의 의미 있는 관계 및 규칙, 즉 일정한 패턴을 추출하는 것이다. 데이터 마이닝 알고리즘은 데이터 세트에서 일정한 관계를 식별하도록 설계가 되어있고 이 알고리즘이 정형 데이터 및 비정형 데이터 세트에서 자동으로 패턴을 식별하게 된다.

② 기계 학습(Machine Learning)

데이터를 바탕으로 학습이 가능한 소프트웨어를 사용하여 컴퓨터에게 학습 능력을 부여하는 것을 말한다. 이 방법의 핵심은 기존에 학습한 알고리즘을 기반으로 미래를 '예측'하는 것이다.

③ 딥러닝(Deep Learning)

기계 학습의 한 부분으로, 인간의 신경망(Neural Network)과 유사하게 다수의 노드(Node)로 구성된 인공신경망을 구축하여 인간의 뇌와 유사한 방식으로 문제를 해결하도록 하는 분석이다.



[그림 3-19] 인공지능, 머신러닝, 딥러닝의 관계

(출처 : Live LG, <https://live.lge.co.kr/>)

① 사회 연결망 분석(Social Network Analysis)

각 개인을 노드(node), 각 관계를 연결선(Ties)으로 표현하여 집단과 집단, 집단과 사람 사이에 어떤 연결 관계가 있는지 파악하는 것이다.

② 데이터 시각화(Data Visualization)

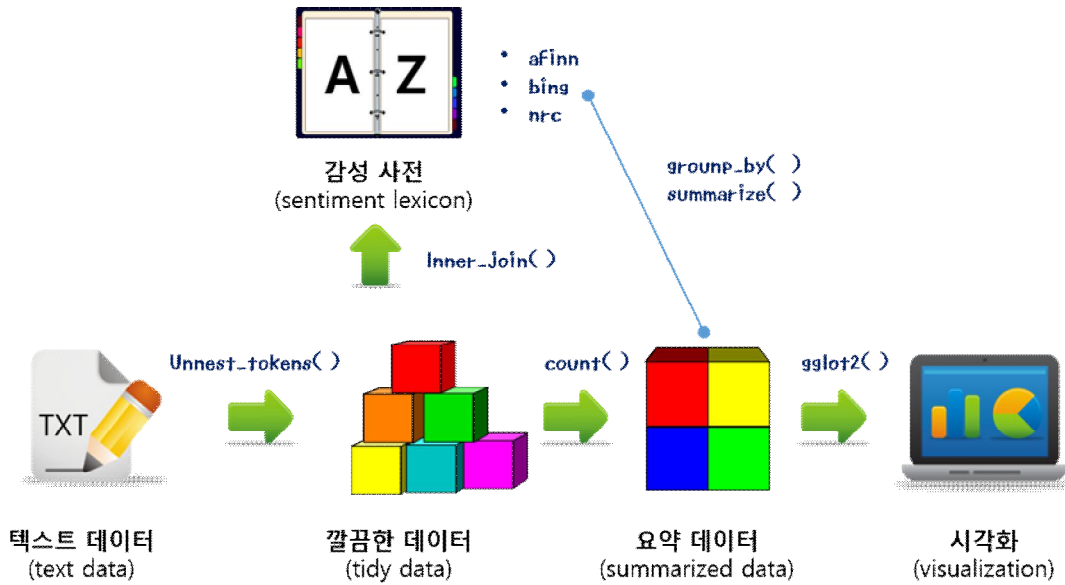
분석으로 도출한 결과를 차트 및 그래프 등의 시각적인 자료로 나타내는 방법이다.

③ 통계 분석(Statistic Analysis)

둘 이상의 변수들 사이의 관계를 파악하거나 집단 간의 통계적인 차이를 분석한다. 대표적으로 상관 분석, 회귀 분석, 분산 분석 등이 있다.

④ 감성 분석(Sentiment Analysis)

텍스트 데이터에서 감정을 나타내는 속성을 추출하여 패턴화하고, 텍스트에 나타난 감정을 긍정/부정으로 판별하는 분석이다.



[그림 3-20] 감성 분석의 과정

16.5. 시각화 및 활용

가. 시각화의 의미

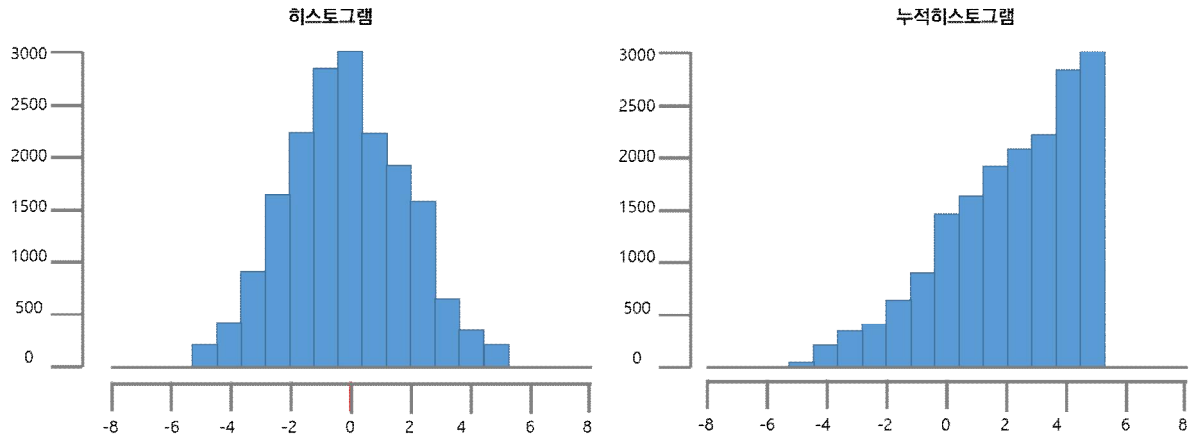
빅데이터 시대에는 다양한 유형의 데이터와 정보가 넘쳐나고 있다. 따라서 조직 구성원 또는 관리자들은 그 중 중요한 정보와 그렇지 않은 정보를 가려낼 수 있는 능력을 가지고 있어야 한다. 데이터 시각화는 트렌드, 시장 수요, 처리 과정 향상을 전체적으로 이해하기 위해 전 세계 여러 조직에서 사용하는 아주 효과적인 도구이다. 시각화 자료들은 의사 결정 과정에 도움을 주는데 예를 들어 차트, 그래프, 지도 등은 시각적인 신호들을 효과적으로 사용함으로써 숫자로 이루어진 표보다 더욱 설득력이 있으며 관리자들은 대규모의 정보를 한눈에 처리하고 파악할 수 있게 된다.

나. 주요 시각화 방법 및 활용

데이터를 시각화하기 전에는 데이터의 유형과 데이터 분석의 목적을 고려해야 한다. 시각화한 자료는 다양한 목적으로 활용이 가능한데, 크게 시간 경과에 따른 데이터의 변화를 알아보거나 전체의 각 구성 요소의 크기를 비교하거나 변수들 사이의 관계 및 분포를 한눈에 쉽게 알아보기 위해 활용할 수 있다. 일반적인 시각화 기법은 시간, 분포, 관계, 비교 그리고 공간 시각화가 있다.

① 시간 시각화

시간의 경과에 따른 하나 혹은 여러 개의 변수 데이터의 변화를 표현해준다. 일정한 기간에 걸친 변화 및 트렌드를 추적하여 시간의 전후 관계를 분석할 때 사용한다. 시간 시각화 방법으로는 막대그래프, 누적막대그래프, 묶은막대그래프, 점그래프, 꺾은선그래프, 계단그래프, 추세선 등이 있다.



[그림 3-21] 히스토그램과 누적히스토그램

② 분포 시각화

특정 변수에 해당하는 값들의 분포를 시각적으로 표현해주어 최대, 최소값 등을 알기 쉽다. 예를 들어 히스토그램, 파이차트(원그래프), 도넛차트, 트리맵 등이 있으며 전체의 관점에서 보면 각 구성 요소가 차지하는 비율을 비교 및 파악하기 쉽다.



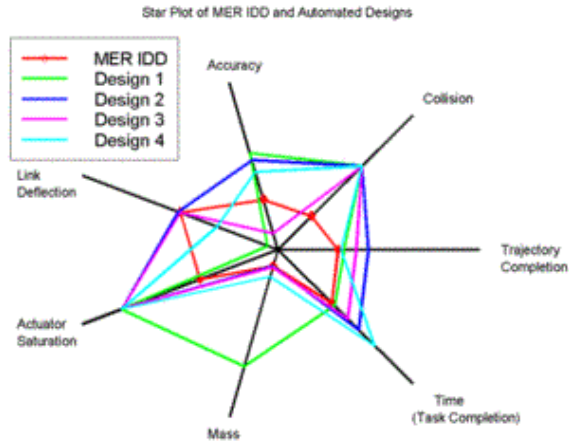
[그림 3-22] 시도별 인구 분포 트리맵

③ 관계 시각화

2개 이상의 변수 데이터 사이의 관계를 표현한다. 특히 히트맵은 복잡한 통계 데이터를 보여줄 때 가장 많이 사용되는 방법으로 따뜻한-차가운 색상의 스펙트럼을 사용해서 데이터의 다른 계층을 보여준다. 대표적으로 산점도(스캐터플롯), 히트맵, 버블차트 등이 있다.

④ 비교 시각화

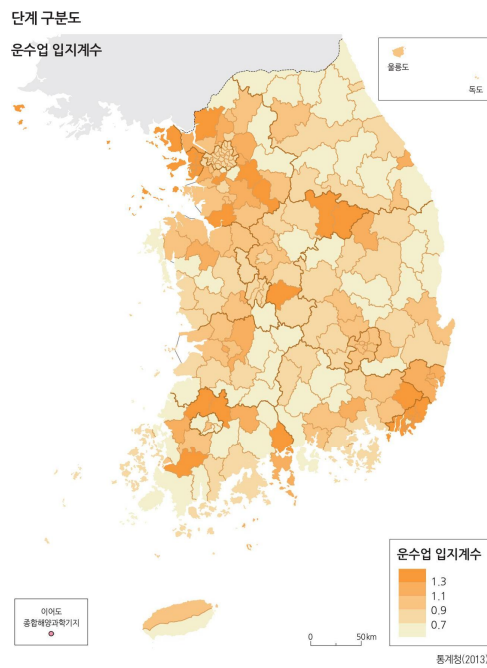
여러 개의 변수 사이의 차이 및 유사성을 나타낸다. 즉 하나의 항목에 대해 여러 차원의 변수를 비교할 수 있다. 예를 들어 히트맵, 스타차트(웹차트,스파이더차트), 평형좌표계, 다차원척도법을 사용할 수 있다.



[그림 3-23] 스타 차트

⑤ 공간 시각화

지리적 위치에 따른 데이터가 어떻게 분포되었는지 나타내는 것으로 실제 지도를 사용하여 데이터를 시각화한다. 대표적으로 지도 맵핑, 단계구분도, 카토그램 등이 있다.



[그림 3-24] 단계 구분도
(출처 : 국토지리정보원)

다. 빅데이터 시각화 도구

빅데이터 시각화를 위한 기본적인 도구로는 Excel, Cvs/JSON, Chart API, D3, Google 등을 사용할 수 있다. 또한, 맵핑을 위한 대표적 도구로 Modest Maps, Polymaps, OpenLayers 등이 있다.

전문적으로는 NodeBox, R, Weka 프로그램을 사용할 수 있다. 그 외 코딩 없이 다양한 시각화가 가능한 도구로는 Tableau, Infogram, ChartBolcks, Datawrapper 등이 있다. 이 중 Tableau가 빅데이터 시각화에 가장 많이 사용되고 있다.

16.6. 데이터 폐기

가. 데이터 폐기 개요

빅데이터, 인공지능, 사물인터넷과 같은 기술을 통해 기업이 가진 데이터의 규모도 점차 커지고 있다. 따라서 기업에 대규모의 데이터를 효율적으로 운영 및 관리하는 능력도 요구되고 있다. 더 이상 사용하지 않거나 데이터명을 확인할 수 없는 데이터를 일명 '다크 데이터'라고도 하는데, 다크 데이터가 증가할수록 데이터를 관리할 경제적인 비용이 증가할 뿐 아니라 전력 소모가 커 환경에도 영향을 미치게 된다.

그러므로 데이터관리를 위한 전력 소비를 줄이는 노력 외에 불필요한 데이터를 삭제하여 데이터의 양을 줄이려는 노력이 필요하다고 할 수 있다.

데이터 폐기는 서비스가 종료되거나 서비스를 이전할 때 데이터 자체를 더 이상 보유하거나 관리하지 않고 다시 사용할 수 없도록 정보를 삭제하는 것을 의미한다. 이때 데이터는 복구할 수 없도록 삭제하는 것이 원칙이며 이미 백업이 되어있는 데이터에 대해서도 어떻게 삭제할 것인지에 대한 방안이 구체적으로 마련되어 있어야 한다.

나. 올바른 데이터 폐기

데이터를 삭제하였더라도 데이터가 하드디스크에서 완전히 삭제된 것은 아니다. IT 저장 매체의 데이터를 폐기 방법은 크게 물리적으로 하드웨어를 파괴하는 방식과 소프트웨어 프로그램을 이용하여 데이터를 삭제하는 방식으로 나뉜다.

하드웨어 파괴는 저장매체에 물리적인 힘으로 구멍을 내 내부를 파괴하거나 저장매체의 기록과 관련된 자기 기능을 손상하는 방법으로 나뉜다. 소프트웨어 삭제는 전문 프로그램을 사용하여 데이터를 덮어쓰기를 하는 방법이나 하드웨어를 포맷하는 방법 등이 있다.

[표 3-6] 데이터 폐기 방법

| 구분 | 방법 | 데이터 폐기 방법 |
|---------|------------------------------|---|
| 하드웨어 파괴 | 물리적 파괴 및 분쇄 | 물리적인 힘으로 구멍을 내어 내부를 파괴 |
| | 자기적 삭제 (magnetic erasing) | 저장 매체의 기록과 관련된 자기 기능을 손상 |
| 데이터 삭제 | 덮어쓰기 (override) | 기존에 저장된 데이터 공간에 다른 데이터를 덮어쓰워 기존의 데이터를 복원하지 못 하게 하는 방법 |
| | 로레벨 포맷 (low level format) | 덮어쓰기 기법과 비슷하지만 하드 디스크를 포맷하는 방법. 매우 오래 걸리며 여러 번 수행해야 한다. |

참고 문헌

- 김강원. 『실무로 배우는 빅데이터 기술』, 위키북스 (2020),p16-20.
- 이윤배. 『개념에 충실한 데이터베이스론』, 휴먼사이언스 2016),p558-561.
- 이훈. 『사물인터넷과 생태계』,한티미디어 (2020),p179-187.
- 윤종식. 『2022 ADsP데이터분석 준전문가』, DATADU (2021)
- 윤희성, 최건우, 황수진, '빅데이터 분석은 해운에서 어떻게 활용되는가?', 한국해양수산개발원, 2018.07.
- 해양수산개발원-해양수산빅데이터 거래소 (<https://www.bigdata-sea.kr/>)
- 황석형, 김응희, 김민경. 『데이터사이언스 입문, 인공지능시대의 전문서』, 인피니티북스 (2020),p14-25.
- Lee, M. Y., & Choe, W. (2012). 빅데이터 분석을 위한 빅데이터 처리 기술 동향. Korea Information Processing Society Review, 19(2), 20-28.
- Zielinski, J. S. (2018) Does Smart Grid need new Informatics Tools?.Przegląd Elektrotechniczny,4, pp. 30-33.

VII. 통신기술 기본

목포해양대학교 **홍성화** 교수



17. 정보통신 기초

17.1. 정보통신의 개념

정보통신은 정보와 통신 기술이 결합되어 서로 다른 지점 또는 장치 간에 데이터/정보/지식을 전송하고 공유하는 것을 말한다. 우선, 정보, 데이터, 그리고 지식은 서로 다른 개념이지만, 밀접한 관련이 있다. 각각의 개념은 다음과 같이 정의될 수 있다.

① 데이터(Data)

단순히 수집된 사실과 숫자들의 집합으로, 아직 분석되거나 가공되지 않은 원시 자료를 뜻하며, 주로 사실(Fact) 이나 상황(Situation)에서 관찰과 측정으로 발생

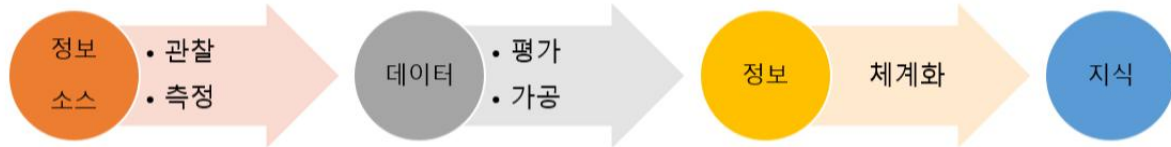
② 정보(Information)

데이터를 처리하고 분석한 결과물로, 의사결정을 위한 데이터를 지칭하며 지식을 전달하는데 사용. 즉, 데이터로부터 의미 있는 패턴, 트렌드, 인사이트 등을 도출하여 가치를 창출

③ 지식(Knowledge)

정보를 분석, 해석하고 응용하여 얻은, 이해하고 활용할 수 있는 것으로, 지식은 정보를 이해하고 활용하여 문제를 해결하고 미래를 예측하는 데 사용한다. 이러한 개념들은 서로 유기적으로 연결되어 있다. 데이터를 수집하고 분석하여 정보를 도출하고, 이 정보를 이해하고 응용하여 지식을 얻으며, 지식은 다시 데이터를 수집하고 분석하는데 사용될 수 있다. 요약하면, 데이터는 정보를 생성하는 원시 자료이며, 정보는 데이터의 분석과 처리를 통해 생성되고, 지식은 정보를 이해하고 응용하

여 얻은 것이다. 이러한 관계를 이해하고 이용하여 데이터를 적극적으로 수집하고 분석하여 정보와 지식을 창출하는 것이 중요하다.

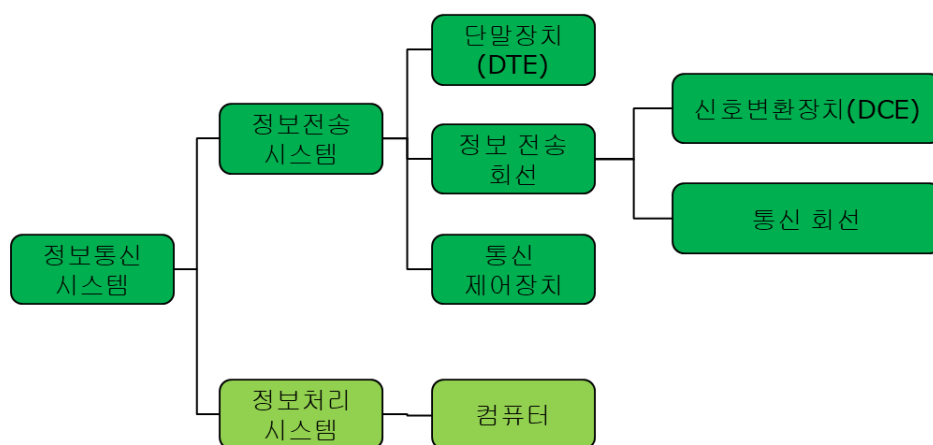


[그림 3-25] 데이터, 정보, 지식 간의 상관 관계

정보통신 기술은 컴퓨터, 인터넷, 통신망 등 다양한 기술을 포함하며, 이러한 기술을 통해 사람들은 음성, 영상, 데이터 등 다양한 형태의 정보를 교환하고 활용할 수 있다. 즉, 정보 통신 기술은 정보 기술 (Information Technology)과 통신 기술(Communication Technology)의 합성어로서, 주로 ICT (Information and Communication Technology)로 언급된다.

이러한 다양한 정보통신 기술을 실제 하드웨어와 소프트웨어를 이용하여 구현한 정보통신 시스템은 정보를 전달하기 위해 구축된 정보전송 시스템과 정보를 저장, 처리 및 관리하는 정보처리 시스템으로 구성된다고 할 수 있다.

따라서, 정보 통신은 정보 기기의 하드웨어, 운영 및 정보 관리에 필요한 소프트웨어 기술, 이들 기술을 이용하여 정보를 수집, 생산, 가공, 보존, 전달, 활용하는 모든 방법으로서, 필요한 자료를 수집하고, 가공하고, 재창출하기 위해서 필요한 과학적 지식과 관련된 학문으로, ICT 기술의 주요 구성요소로는 컴퓨터와 통신망이 있다. 컴퓨터는 정보를 처리하고 저장하기 위한 기기이며, 통신망은 컴퓨터들 간의 연결을 가능하게 해주는 매체이다.



[그림 3-26] 정보통신 시스템의 구성요소

※ 정보통신 기술의 특징

① 신속성

빠른 속도로 정보를 전송하고 처리함으로써, 사용자는 실시간으로 정보를 주고받을 수 있으며, 빠른 응답시간을 보장받을 수 있음

② 확장성:

정보통신 기술은 시스템의 크기와 복잡도를 늘릴 수 있는 확장성을 가지고 있음

③ 유연성

정보통신 기술은 다양한 종류의 기기와 네트워크에 적용될 수 있음

④ 보안성

개인 정보와 기업 비밀 정보 등을 안전하게 보호하고, 해커와 같은 악의적인 공격으로부터 방어

⑤ 대규모 처리

대규모의 데이터를 처리하기 위해 분산 처리 기술이나 클라우드 컴퓨팅 기술 등이 사용

⑥ 개방성

다양한 플랫폼에서 서비스를 개발하거나 사용할 수 있음

17.2. 정보전송 기초

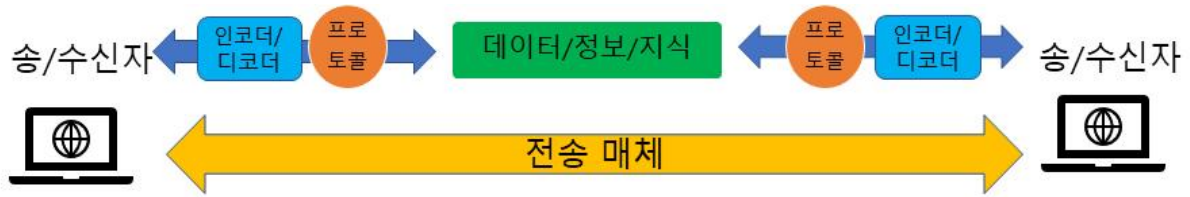
정보소스에서 얻은 데이터/정보/지식을 송신자가 인코딩하고, 채널을 통해 전송하며, 수신자가 정보를 디코딩하여 수신하는 과정으로 이루어지는 것을 정보통신 기술 중의 정보 전송(Information transmission) 과정이라고 한다. 정보 전송은 다양한 방식의 유무선 매체를 통해 이루어질 수 있다. 가장 일반적인 방식은 유선 또는 무선 통신망을 통해 디지털 혹은 아날로그 신호를 통한 전송 방법이다.

아날로그 신호는 연속적인 값으로 이루어진 신호로서 물리적인 크기의 변화에 따라 전압, 전류, 주파수 등의 연속적인 값이 변화한다. 예를 들어, 소리는 아날로그 신호로 전달됩니다. 소리의 높낮이와 크기가 연속적인 값으로 변화하기 때문이다.

반면, 디지털 신호는 0과 1로 이루어진 이산적인 신호로서 아날로그 신호를 디지털 신호로 변환하면서

값을 샘플링하고 양자화하는 과정을 거쳐 디지털 신호로 변화한다. 이 신호는 노이즈나 왜곡에 강하며, 또한 오류를 검출하고 수정하는 기술이 존재하여 신호를 검증할 수 있다.

또한, 디지털 신호는 아날로그 신호와 달리, 주파수 대역폭이 좁아서 더 많은 정보를 전송할 수 있다. 따라서 디지털 통신은 더 많은 정보를 빠르고 안정적으로 전송할 수 있으나, 디지털 신호를 만들기 위해서는 아날로그 신호를 디지털 신호로 변환하는 과정이 필요하며, 이를 위해 필요한 장비와 기술이 필요하다.



[그림 3-27] 정보 전송의 구성요소

정보 전송은 다양한 구성요소로 이루어져 있으며 주요 구성요소는 다음과 같다.

① 송신자(Sender)

정보를 전송하는 역할

② 수신자(Receiver)

정보를 받는 역할

③ 전송 매체(Medium/Channel)

정보를 전송하는 통로

④ 인코더(Encoder)

송신자가 만든 정보를 채널에 전송하기 위해 적절한 형태로 변환하는 역할을 하며, 소리를 전송하기 위해서는 아날로그 신호를 디지털 신호로 변환

⑤ 디코더(Decoder)

수신자가 받은 정보를 해석하고 처리하기 위해 송신자가 보낸 정보를 이해할 수 있는 형태로 변환하는 역할을 하며 인코더와 반대로 동작

⑥ 프로토콜(Protocol)

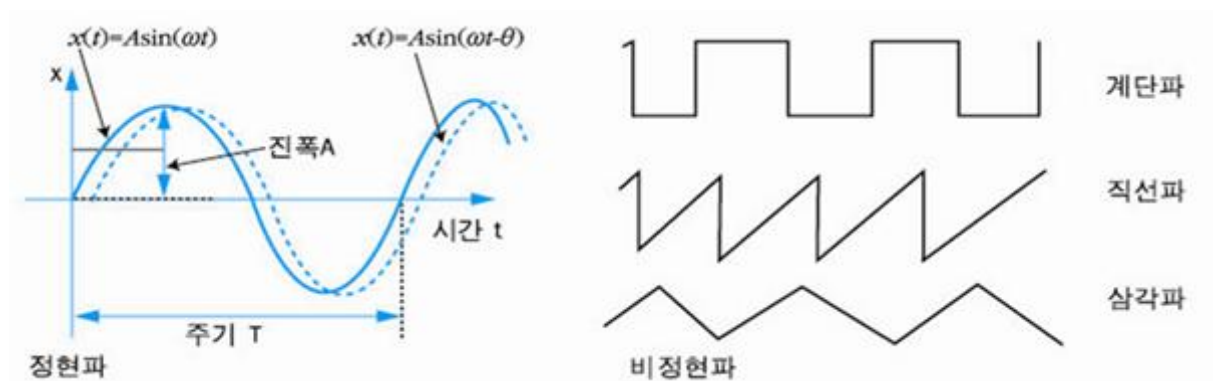
정보를 전송하기 위해 사용되는 규칙과 규약으로, 인터넷에서는 TCP/IP 프로토콜이 사용되며, 이를 통해 정보가 안전하고 정확하게 전송

(1) 아날로그 신호

아날로그 신호는 연속적인 값을 가지는 신호로서, 시간이 지나면서 값이 연속적으로 변화합니다. 이러한 신호는 물리적인 현상(예: 소리, 빛, 전기 신호 등)으로부터 생성될 수 있으며, 주로 연속적인 파형으로 표현됩니다. 아날로그 신호는 아날로그 값(Analog Value)이라는 연속적인 값을 가지며, 이 값은 일반적으로 전압, 전류, 음압 등으로 측정됩니다. 이러한 값은 시간에 따라 변화하며, 일반적으로 연속적인 파형으로 표현됩니다.

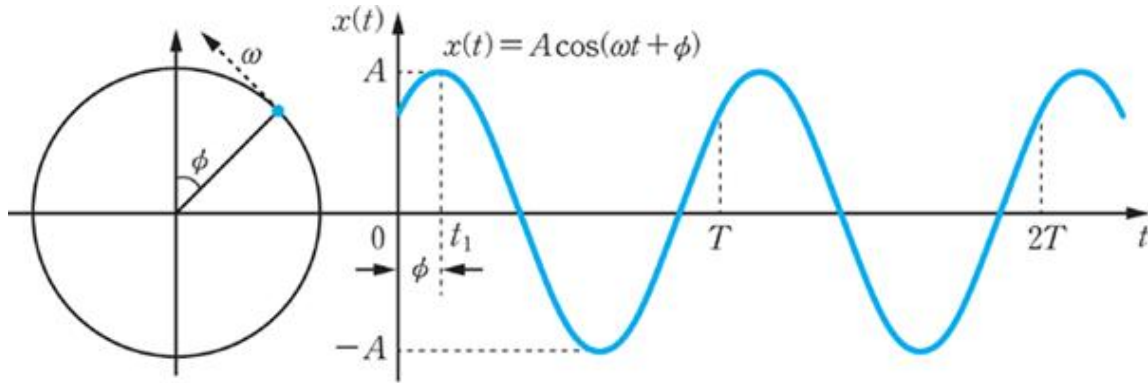
이러한 파형은 주기적이거나 비주기적일 수 있으며, 주파수, 진폭, 위상 등의 특성을 가집니다.

아날로그 신호는 수치적으로 표현될 수 있으며, 이는 아날로그-디지털 변환(ADC)을 통해 이루어 집니다. ADC는 아날로그 신호를 샘플링하여 디지털 신호로 변환하는 장치입니다. 이러한 디지털 신호는 이산적인 값을 가지며, 일반적으로 0과 1로 표현됩니다.



[그림 3-28] 정현파와 비정현파

아날로그 신호는 일반적으로 주기 신호와 비주기 신호로 분류되는데, 주기 신호는 정현파와 비정현파로 분류된다. 비정현파에는 계단파, 직선파, 삼각파 등이 있는데 주로 컴퓨터 내부 클럭 파형이다.



[그림 3-29] 정현파의 발생

정현파는 사인 함수 또는 코사인 함수의 형태로 나타나는 주기적인 아날로그 신호입니다. 일반적으로 정현파는 다음과 같은 수식으로 표현됩니다.

사인 함수를 사용한 정현파:

$$x(t) = A \sin(2\pi ft + \theta)$$

코사인 함수를 사용한 정현파:

$$x(t) = A \cos(2\pi ft + \theta)$$

여기서, $x(t)$ 는 시간 t 에서의 정현파의 값(아날로그 값)을 나타내며, A 는 진폭(amplitude)을 나타냅니다. f 는 정현파의 주파수(frequency)를 나타내며, 단위는 Hz(헤르츠)입니다. θ 는 위상(phase)을 나타내며, 단위는 라디안(radian)입니다. 주기 T 는 주파수 f 와의 관계에서 $T = 1/f$ 로 계산됩니다.

사인 함수와 코사인 함수는 둘 다 원의 호를 나타내는 삼각함수 중 하나입니다. 정현파는 이러한 함수를 사용하여 시간에 따른 파동 모양을 나타냅니다. 정현파는 아날로그 신호 중에서도 가장 기본적인 형태로, 실제로 많은 신호들은 여러 개의 정현파의 합으로 구성됩니다.

모든 아날로그 신호는 이론적으로 정현파들의 합으로 나타낼 수 있습니다. 이를 푸리에(Fourier) 변환이라고 합니다. 푸리에 변환은 임의의 아날로그 신호를 주파수 영역으로 변환해주는 수학적 기법입니다. 이를 통해 아날로그 신호를 구성하는 다양한 주파수의 정현파 성분들로 분해할 수 있습니다. 이러한 성분들을 구성하는 주파수와 진폭 정보를 알고 있다면, 아날로그 신호를 다시 재구성할 수 있습니다.

따라서, 모든 아날로그 신호는 이론적으로 주파수 영역에서 정현파들의 합으로 분해될 수 있으므로, 정현파들의 조합으로 표현될 수 있습니다.

하지만, 이를 구하는 것은 실제로는 복잡하고 수학적으로 어려울 수 있으며, 컴퓨터를 이용하여 수치적으로 근사값을 구하는 것이 일반적입니다.

(2) 디지털 신호

디지털 신호는 이산적인 값을 가지는 신호로서, 디지털 정보처리를 위해 사용됩니다. 이산적인 값이란, 연속적인 값이 아닌, 몇 가지 불연속적인 값을 가지는 것을 의미합니다.

디지털 신호는 0과 1로 이루어진 비트(bit)의 집합으로 표현됩니다. 이진수(binary)로 표현하는 것이기 때문에, 이진수에서 사용되는 2진법의 개념을 이해해야 합니다. 2진법이란, 0과 1 두 개의 숫자만을 사용하여 숫자를 표현하는 방법입니다.

예를 들어, 8비트의 디지털 신호가 있다면, 이는 0과 1로 이루어진 8자리의 이진수(binary)로 표현됩니다. 예를 들어, "01101000"은 8비트의 디지털 신호로서, 이진수에서는 104를 나타냅니다.

디지털 신호의 정의화 표현 방식에는 여러 가지가 있지만, 가장 대표적인 방식은 펄스 코드 변조(Pulse Code Modulation, PCM)입니다. PCM은 아날로그 신호를 디지털 신호로 변환하는 방법 중 하나입니다.

(3) 아날로그와 디지털 전송

• 아날로그 전송

- 장거리 전송 이용
- 감쇠 현상(거리가 멀수록 자주 발생)
- 왜곡 현상이 발생하기 때문에 감쇠 신호 증폭을 위한 증폭기(Amplifier) 필요

• 디지털 전송

- 제한된 거리에서만 사용 => 현재는 장거리 전송도 가능함
- 감쇠 현상이 발생하면 심각한 오류임
- 감쇠 신호 재생을 위한 리피터(Repeater) 사용
- 데이터 무결성 보장
- 광대역을 위한 고도의 다중화 기법을 이용한 값싼 전송 용량의 활용
- 보안을 위한 암호화(encryption) 기법 사용 가능
- 모든 종류의 신호를 동일한 신호로 통합함으로써 경제적임

- 전송 신호의 손상

- 손상의 원인

신호가 전송 매체를 따라 전송될 때 신호의 크기가 감소되고, 모양이 일그러지며(왜곡), 잡음이 침입하기 때문

- 신호 감쇠(signal attenuation) :

저항이나 임피던스의 영향으로 전송 거리에 비례하며 점점 약해지는 현상

감쇠의 정도는 단위 거리에 따라 로그 함수의 비율(dB로 표시)

아날로그 전송에서는 일정한 거리마다 증폭기 설치

디지털 전송에서는 리피터 설치

- 지연 왜곡(delay distortion) : 전달되는 신호의 전파 속도가 주파수에 따라 차이나는 것

감쇠의 정도가 균일하지 않아 발생

심벌 간 간섭(inter-symbol interference) 현상이 일어남

등화기(equalizer) 로 해결

- 잡음(noise) :

전송 시스템 및 전송 매체에서 발생하여 전송 신호에 추가되는 불필요한 신호

통신 시스템의 성능을 떨어뜨리는 주요 원인

열 잡음(thermal noise, white noise), 상호 변조 잡음(inter-modulation noise), 누화 잡음(cross talk noise), 충격 잡음(impulse noise)

17.3. 정보전송 방식

정보 전송 방식은 정보를 한 장소에서 다른 장소로 전송하기 위해 사용되는 방법을 말합니다. 이는 정보를 전달하기 위한 매체, 프로토콜, 기술 등을 사용하여 이루어지며, 네트워크를 통해 이루어질 수도 있습니다. 정보 전송 방식에는 직렬/병렬 방식, 동기/비동기 방식, 통신 방식에 따라 여러 유형으로 나누어지며, 현재 아날로그/디지털 통신에서 주로 사용되는데 본 교재에서는 주로 데이터 통신을 중심으로 언급한다.

(1) 직렬/병렬 전송

직렬 전송과 병렬 전송은 정보를 전송하는 방식 중에서 가장 기본적인 두 가지 방식입니다.

(가) 직렬 통신

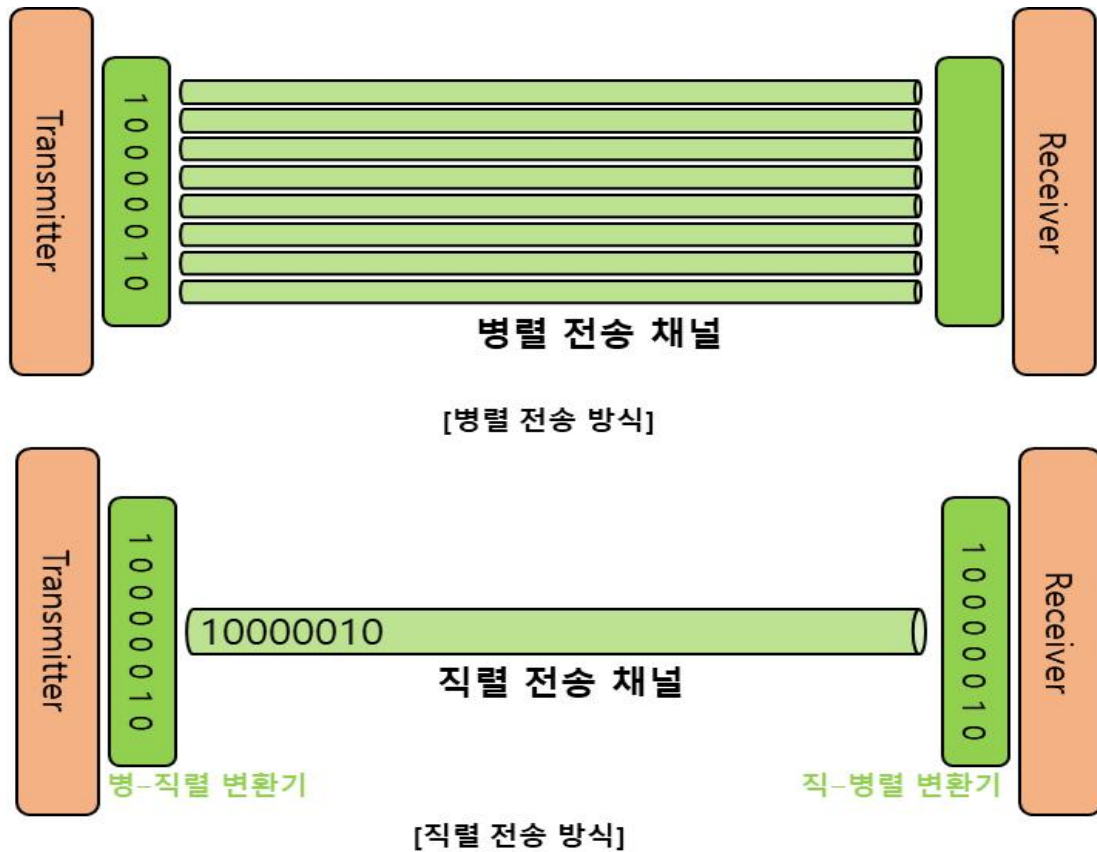
- ① 한 번에 한 비트씩 데이터를 전송하는 방식
- ② 케이블의 길이가 길어질수록 속도 저하가 발생하는데, 이를 해결하기 위해 다양한 기술(예: 레이저 전송, 증폭기 등)을 사용할 수 있음
- ③ 선이 적게 필요하므로 복잡한 연결이 필요하지 않고, 케이블이 적은 수의 데이터 전송에 적합함
- ④ 오류 검사와 수정에 용이함

(나) 병렬 통신

- ① 한 번에 여러 비트를 전송하는 방식으로 한 문자를 구성하는 각 비트를 각각의 데이터 선을 통해 한꺼번에 전달하는 방식
- ② 직렬 통신에 비해 더 빠른 전송 속도를 제공함
- ③ 다수의 데이터를 전송할 때 유리하며, 대용량 데이터를 전송할 때 성능이 우수함
- ④ 케이블의 길이가 길어질수록 전송 속도가 느려짐
- ⑤ 전송 중간에 하나의 비트가 손상되면 전체 전송이 실패할 수 있음

따라서, 직렬 통신과 병렬 통신은 각각의 장단점이 있으며, 특정 상황에 따라 선택해야 합니다. 짧은 거리에서 대용량 데이터를 전송하는 경우에는 병렬 통신이 유리하며, 긴 거리에서 속도가 느리더라도 안정적으로 데이터를 전송해야 하는 경우에는 직렬 통신이 유리합니다.

직렬 전송은 전송 매체가 제한적이거나 전송 속도가 느려도 상관없는 경우에 사용되며, 병렬 전송은 전송 속도가 빠르고 많은 양의 데이터를 전송해야 하는 경우에 사용됩니다. 예를 들어, 하드 디스크나 메모리 같은 컴퓨터 내부의 데이터 전송에서는 병렬 전송이 사용되고, 외부에서 데이터를 전송할 때는 직렬 전송이 사용됩니다.



[그림 3-30] 직렬/병렬 전송

(2) 동기/비동기 방식

(가) 동기 방식

데이터 통신에서의 동기방식(synchronous communication)은 데이터를 전송하기 전에 데이터 송수신 측 간에 타이밍 신호를 주고 받아 동기화하는 방식입니다.

이 방식은 전송하는 데이터를 블록(block) 단위로 전송하며, 블록 내의 모든 데이터가 전송된 후에 동기화 신호가 전송됩니다.

동기방식은 다수의 데이터를 전송할 때 효과적입니다. 여러 비트로 이루어진 데이터를 한 번에 전송하기 때문에 전송 속도가 빠르며, 전송 과정에서 생길 수 있는 오류를 검출하기 쉽습니다.

또한, 전송하는 측과 수신하는 측의 속도 차이가 적은 경우에 유리합니다.

동기방식에서는 송신 측과 수신 측 간에 데이터 전송에 사용할 시간을 미리 정해놓고 사용합니다. 이 때, 송신 측은 데이터를 전송하는 동안 시간을 맞추기 위해 수신 측으로부터 지속적으로 동기화 신호를 받습니다. 수신 측은 이 동기화 신호를 기반으로 데이터를 받을 준비를 합니다.

동기화를 위한 신호를 만들어 전송하는 방법으로는 클럭(clock) 신호를 이용하는 방식과 스타트-스톱(start-stop) 신호를 이용하는 방식이 있습니다.

클럭 신호 방식에서는 시계 신호를 이용하여 데이터 전송 시간을 정하고, 스타트-스톱 신호 방식에서는 시작과 끝을 나타내는 신호를 이용합니다.

하지만 동기방식의 경우, 블록 단위로 데이터를 전송하는 특성상, 전송하는 블록의 크기가 작은 경우 전송 속도가 저하될 수 있습니다.

또한, 송신 측과 수신 측의 속도 차이가 크면 동기화가 어렵고, 전송 중간에 손상된 데이터를 복구하기 어렵습니다. 따라서, 이러한 단점을 보완하기 위해 비동기 방식을 사용하기도 합니다.

(나) 비동기 방식

데이터 통신에서의 비동기방식(asynchronous communication)은 데이터를 전송할 때 타이밍 신호를 사용하지 않고, 데이터의 시작과 끝을 특정 문자나 신호로 나타내어 전송하는 방식입니다.

이 방식은 데이터를 문자 단위로 전송하기 때문에 전송 속도가 느리지만, 전송하는 데이터의 크기에 제한이 없기 때문에 큰 데이터를 전송할 때 유용합니다.

비동기방식에서는 데이터 전송 시간이 일정하지 않으므로, 수신 측에서는 전송되는 데이터의 시작과 끝을 인식하기 위해 특정한 문자나 신호를 기준으로 데이터를 구분합니다.

이를 위해 비트 시간 간격을 이용하여 데이터를 구분하기도 합니다. 이러한 방식으로 데이터를 전송하기 때문에, 전송 중간에 손상된 데이터를 복구하기 위해 오류 검출 및 정정 기능이 필요합니다.

비동기방식은 전송하는 데이터의 크기에 제한이 없고, 데이터 전송 시간이 일정하지 않아도 되므로 데이터를 전송하는데 있어서 자유도가 높습니다.

또한, 데이터 전송 시간이 일정하지 않기 때문에 송신 측과 수신 측 간의 속도 차이가 크더라도 전송이 가능합니다.

하지만, 데이터 전송 속도가 느리고 오류 검출 및 정정 기능이 필요하기 때문에 전송 속도가 빠른 동기방식과는 달리 효율성이 떨어질 수 있습니다.

데이터 통신에서는 동기방식과 비동기방식을 적절히 조합하여 사용하기도 합니다. 예를 들어, 전송할 데이터의 크기가 작고, 전송 속도가 빠른 동기방식을 사용하여 전송 속도를 높이고, 전송할 데이터의 크기가 크거나 전송 속도가 느린 경우에는 비동기방식을 사용하여 전송합니다.

이러한 방식을 혼합 동기방식(mixed synchronous/asynchronous communication)이라고 합니다.

(3) 통신 방식

(가) 단방향 통신(Simplex Communication)

단방향 통신(Simplex communication)은 데이터를 한 방향으로만 전송하는 통신 방식으로, 데이터를 보내는 쪽과 받는 쪽이 명확하게 구분되어 있으며, 데이터 전송 방향이 고정되어 있으며, 데이터를 받는 쪽에서는 보내는 쪽으로 응답을 보내지 않습니다.

단방향 통신은 데이터를 전송하는 쪽과 받는 쪽이 역할이 분명하게 구분되는 있는데, 예를 들어, 라디오 방송, 텔레비전 방송, 기상 위성 등에서 데이터를 한 방향으로 전송하는 경우가 대표적이다. 이러한 경우, 데이터 전송 방향이 고정되어 있기 때문에, 전송 속도가 빠르고, 데이터 손실이나 오류가 발생할 확률이 적습니다. 하지만, 단방향 통신에서는 데이터를 받는 쪽에서 데이터를 보내는 쪽으로 응답을 보낼 수 없기 때문에, 통신이 필요한 두 대의 장비가 서로 대화를 나눌 수 없습니다.

단방향 통신은 데이터 전송 방향이 고정되어 있기 때문에, 전송 속도가 빠르고, 데이터 손실이나 오류가 적은 등의 장점이 있지만, 데이터를 받는 쪽에서 응답을 보낼 수 없고, 데이터 손상이나 오류가 발생한 경우에 대처할 수 있는 기능이 없기 때문에, 통신 방식을 선택할 때는 이러한 점을 고려해야 합니다.

(나) 반이중 통신(Half-Duplex Communication)

반이중 통신(Half-Duplex Communication)은 데이터를 양방향으로 전송할 수 있는 통신 방식 중 하나입니다. 하지만 데이터를 전송하는 쪽과 받는 쪽은 동시에 데이터를 전송할 수 없으며, 한 번에 한 방향으로만 데이터를 전송할 수 있습니다.

즉, 한쪽에서 데이터를 전송하면 다른 쪽에서는 동시에 데이터를 받을 수 없고, 그 반대의 경우도 마찬가지입니다. 반이중 통신은 전송 방향이 번갈아가며 바뀌는 방식으로 동작합니다. 먼저, 한 쪽에서 데이터를 전송할 때, 다른 쪽은 데이터를 수신하기 위해 대기 상태로 들어갑니다.

데이터를 전송하는 쪽이 데이터를 모두 전송하면, 다른 쪽은 이를 수신하고, 그 다음에 데이터를 전송하기 위해 다시 데이터를 전송하는 쪽에게 전송 권한을 요청합니다. 이러한 방식으로 데이터 전송 방향을 계속해서 바꾸면서 양방향 통신을 수행합니다.

반이중 통신은 전송 방향이 번갈아가며 바뀌기 때문에, 동시에 양방향으로 데이터를 전송할 수 없으며, 데이터를 전송하는 쪽과 받는 쪽의 역할이 분명하게 구분되어야 합니다. 이러한 통신 방식은 트랜시버(transceiver)라는 장비를 사용하여 구현할 수 있으며, 대표적인 예로는 무선 랜(Wi-Fi)이나 블루투스(Bluetooth) 등이 있습니다.

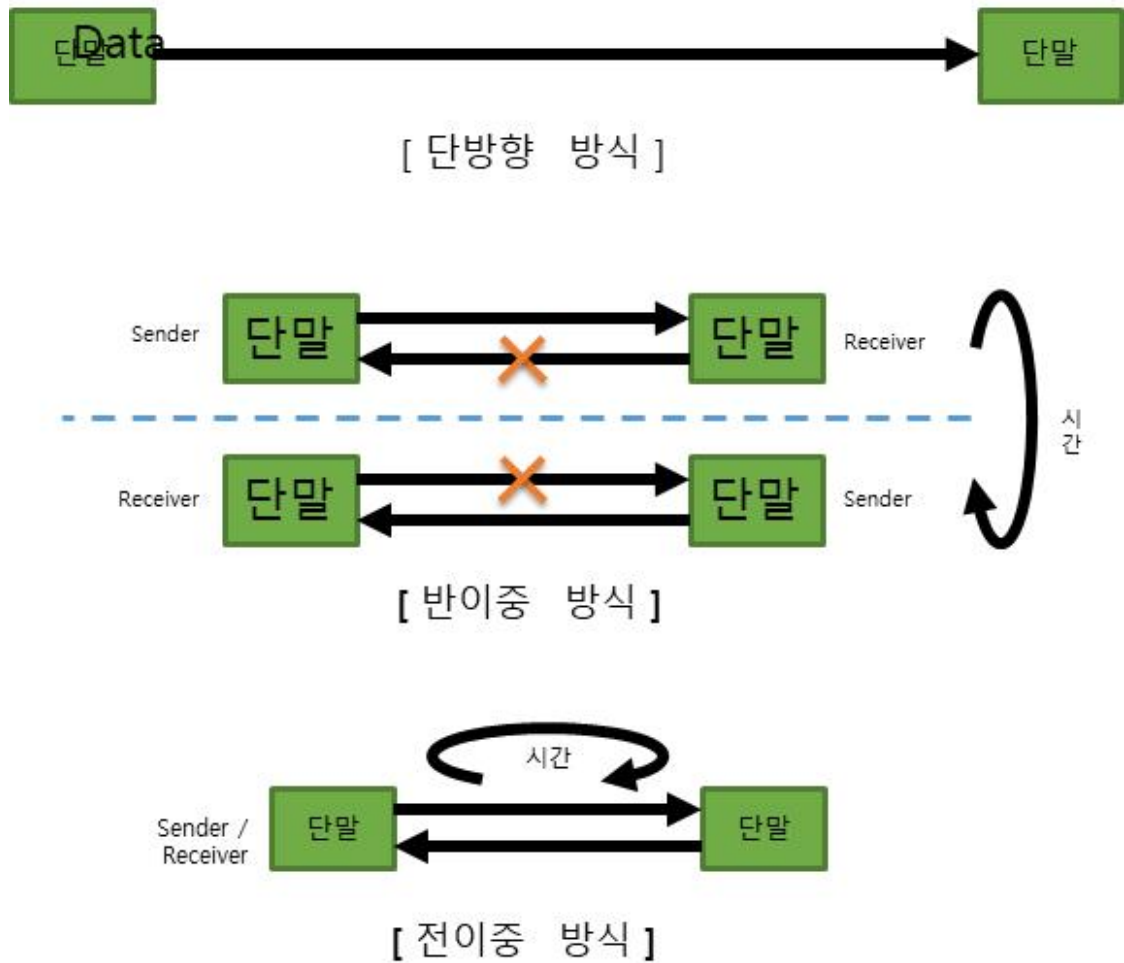
반이중 통신은 전송 방향이 번갈아가며 바뀌기 때문에, 양쪽에서 동시에 데이터를 전송할 수 없는 단점이 있지만, 전송 속도가 빠르고, 두 대의 장비 사이에서 상호작용할 수 있는 기능을 제공하기 때문에, 대부분의 통신 상황에서 유용하게 사용됩니다.

(다) 전이중 통신(Full-Duplex Communication)

전이중 통신은 데이터를 한번에 두 방향으로 동시에 전송하는 통신 방식입니다. 이는 즉, 데이터를 보내는 측과 데이터를 받는 측이 동시에 정보를 주고받을 수 있는 방식입니다.

전이중 통신은 송신자와 수신자가 동시에 정보를 주고받을 수 있기 때문에, 전송 시간이 줄어들어 통신 속도를 높일 수 있습니다. 예를 들어, 전화 통화에서 전화를 걸어 상대방과 이야기할 때 동시에 상대방의 음성을 듣고, 내 음성을 상대방이 들을 수 있습니다. 마찬가지로 인터넷에서 파일을 다운로드하면서 동시에 업로드도 할 수 있습니다.

전이중 통신은 네트워크, 무선 통신 등 다양한 분야에서 사용되며, 데이터를 빠르고 효율적으로 전송하는 데 매우 유용합니다. 전이중 통신의 장점은 양방향으로 데이터를 동시에 전송할 수 있기 때문에 통신의 효율성이 높아진다는 것입니다. 또한, 전이중 통신은 반이중 통신(Half-Duplex Communication)보다 더 안정적이고, 속도도 더 빠릅니다. 전이중 통신은 전송하는 데이터 양이 많을 때 비용이 높아지는 단점이 있습니다. 또한, 전이중 통신을 위해서는 복잡한 회선 구성과 더 많은 대역폭이 필요하기 때문에 시스템 구성이 복잡해질 수 있습니다.



[그림 3-31] 회선 이용 방식

(4) 캐스트 방식

캐스트(Cast) 방식은, 송신자가 특정 그룹에게 데이터를 전송하는 방식입니다. 이 방식은 전체 그룹에게 데이터를 브로드캐스트하는 방식과는 다르게, 특정 그룹에만 데이터를 전송하는 방식입니다.

(가) 유니 캐스트

유니캐스트(Unicast)는 데이터 전송에서 한 송신자가 한 수신자에게만 데이터를 전송하는 방식입니다. 이 방식은 가장 기본적인 데이터 전송 방식으로, 송신자가 특정한 목적지 주소로 데이터를 전송하는 방식입니다.

유니캐스트 방식은 네트워크에서 특정 호스트에게 데이터를 전송할 때 사용되며, 전송 중 오류가 발생할 경우 해당 호스트에 대해서만 재전송을 수행하면 됩니다. 이 방식은 다른 호스트들에게는 영향을 미치지 않으므로, 전송 시간을 절약할 수 있습니다.

유니캐스트 방식은 일반적으로 웹 브라우저에서 서버로의 HTTP 요청, 이메일 등에서 사용됩니다. 이와 달리 멀티캐스트나 브로드캐스트 방식은 전체 또는 그룹의 호스트에게 데이터를 전송하므로, 네트워크 대역폭을 많이 사용하게 됩니다.

(나) 브로드 캐스트

브로드캐스트(Broadcast)는 데이터 전송에서 네트워크 상의 모든 호스트에게 데이터를 전송하는 방식입니다. 즉, 송신자가 특정한 목적지 주소 없이 데이터를 네트워크 상의 모든 호스트에게 보내는 방식입니다. 브로드캐스트는 특별한 상황에서 사용됩니다.

예를 들어, DHCP(Dynamic Host Configuration Protocol)를 사용하여 IP 주소를 동적으로 할당하는 경우, DHCP 서버는 네트워크 상의 모든 호스트에게 IP 주소를 할당하기 위해 브로드캐스트 방식을 사용합니다. 또한, 네트워크 장비의 상태를 모니터링하기 위해 브로드캐스트 메시지를 전송하기도 합니다.

모든 호스트에게 데이터를 전송하기 때문에, 대규모의 네트워크에서는 대역폭을 많이 소모하게 됩니다. 따라서, 이 방식은 일반적으로 작은 네트워크에서만 사용되며, 대규모의 네트워크에서는 멀티캐스트(Multicast)나 유니캐스트(Unicast) 방식이 사용됩니다.

(다) 멀티 캐스트

멀티캐스트(Multicast)는 데이터 전송에서 그룹의 멤버에게 데이터를 전송하는 방식입니다. 즉, 송신자가 그룹의 멤버들에게 데이터를 전송할 수 있는 방식으로, 멀티캐스트 그룹에 속한 모든 호스트에게 데이터를 전송합니다.

멀티캐스트는 브로드캐스트와는 달리, 전체 네트워크 대역폭을 사용하지 않으며, 유니캐스트와는 달리, 그룹에 속한 모든 호스트에게 동시에 데이터를 전송할 수 있습니다. 멀티캐스트 그룹은 IP 주소 범위로 식별되며, IPv4 주소 중 224.0.0.0부터 239.255.255.255 범위 내에서 사용됩니다.

멀티캐스트는 IPTV, 비디오 컨퍼런싱, 멀티미디어 스트리밍, 데이터 센터 및 서버 클러스터, 온라인 게임 등에서 사용됩니다. 이러한 서비스에서는 대규모의 호스트에게 실시간으로 데이터를 전송해야 하므로, 멀티캐스트 방식이 유용합니다.

또한, 멀티캐스트는 대역폭 효율성이 좋으며, 라우팅 효율성도 높아 대규모의 네트워크에서도 사용 가능합니다.

(라) 애니 캐스트

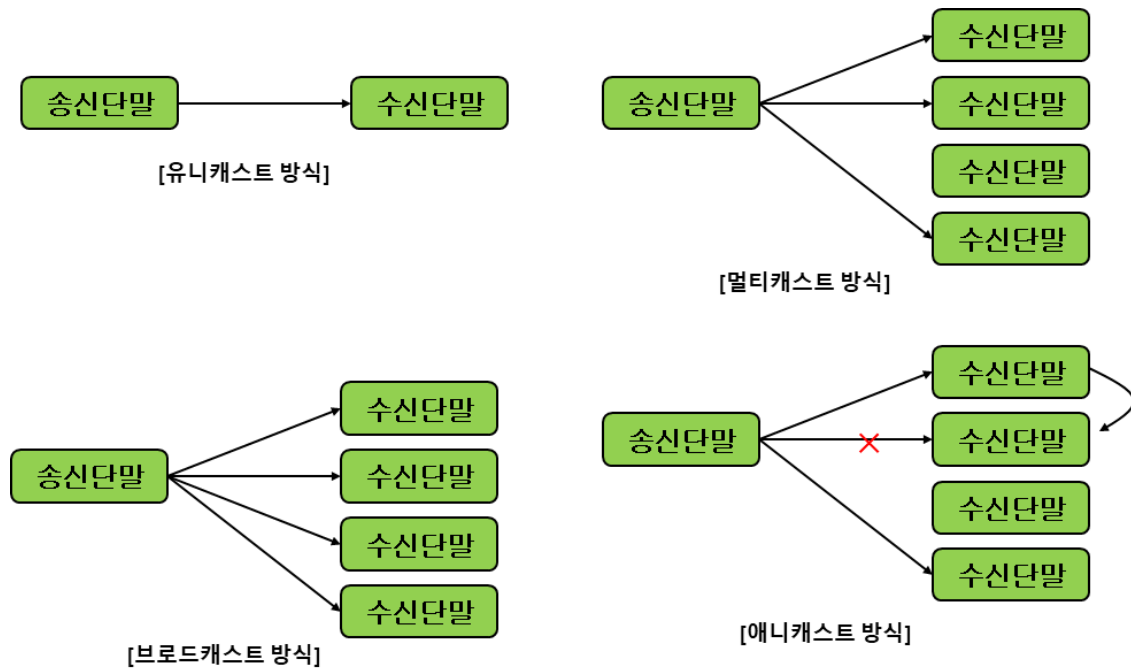
애니캐스트(Anycast)는 IP 네트워크에서 사용되는 전송 방식 중 하나로, 동일한 IP 주소를 사용하는 여러 대의 호스트 중에서 가장 가까운 호스트에게 데이터를 전송하는 방식입니다.

애니캐스트는 멀티캐스트와 유사하지만, 멀티캐스트와는 달리 그룹에 속한 모든 호스트에게 데이터를 전송하는 것이 아니라, 여러 대의 호스트 중에서 가장 가까운 호스트에게만 전송합니다. 이를 통해 로드 밸런싱과 같은 목적으로 사용될 수 있습니다.

애니캐스트는 주로 DNS(Domain Name System) 서비스나 CDN(Content Delivery Network)에서 사용됩니다. DNS 서버는 전 세계적으로 분산되어 있으며, DNS 쿼리 요청에 대한 응답을 처리하는 여러 개의 DNS 서버가 존재합니다.

이 때, 애니캐스트를 사용하여 동일한 IP 주소를 가진 여러 개의 DNS 서버 중에서 가장 가까운 DNS 서버에게 쿼리 요청을 보내, 응답 시간을 최소화할 수 있습니다. 또한, CDN에서는 서버 간의 데이터 복제를 통해 동일한 IP 주소를 사용하는 여러 대의 서버 중에서 가장 가까운 서버에게 클라이언트의 요청을 보내, 콘텐츠 전송 속도를 최적화할 수 있습니다.

애니캐스트는 네트워크 상의 대역폭을 효율적으로 사용할 수 있으며, 서비스 제공자가 서버의 위치를 유연하게 조정할 수 있어 유지 보수와 관리가 쉽습니다.



[그림 3-32] 캐스트 방식

(5) 변조 방식

정보를 전송하기 위해 신호를 변조하는 방식은 다음과 같다.

| [표 3-7] 신호 변환 방식에 따른 변조 방식 | | |
|----------------------------|--------------------|------------------|
| | 신호변환 방식 | 변조방식 |
| 디지털 전송 | 디지털 → 디지털 부호화 | 베이스밴드(Baseband) |
| | 아날로그 → 디지털 부호화 | 펄스 코드 변조(PCM) |
| 아날로그 전송 | 디지털 정보 → 아날로그 부호화 | 브로드밴드(Broadband) |
| | 아날로그 정보 → 아날로그 부호화 | 아날로그 변조 |

(가) 디지털-디지털 부호화 방식

데이터 통신에서 베이스밴드(Baseband)란, 디지털 정보를 직접 디지털 신호로 변환하는 방식을 의미하며 디지털 데이터를 직접 전송하는 방식입니다. 이 방식은 기본적으로 디지털 신호를 전송하고자 하는 비트 스트림으로 표현하고, 이를 직접 케이블이나 무선 매체를 통해 전송합니다.

베이스밴드 방식에서는 아날로그 변조가 필요 없으며, 따라서 변조 과정에서 발생하는 노이즈나 왜곡의 영향을 최소화할 수 있습니다. 이를 통해 전송된 디지털 신호를 정확하게 복원할 수 있습니다.

하지만 베이스밴드 방식은 케이블 또는 무선 매체에서 전송되는 신호가 직접 디지털 신호인 만큼, 전송 거리에 제한이 있습니다. 또한, 이 방식에서는 다른 신호와의 간섭이 발생할 가능성이 높기 때문에, 적절한 에러 감지 및 수정 기능을 구현해야 합니다.

베이스밴드 방식은 주로 LAN, WAN, 케이블 TV 등의 데이터 통신 분야에서 사용됩니다. 이 방식을 사용하는 이유는, 대역폭을 효과적으로 사용할 수 있으며, 데이터의 안정성과 신뢰성을 보장할 수 있기 때문입니다. 또한, 데이터 통신에서 2진수 0과 1비트를 전압값에 대응시키는 방법은 다양합니다. 그 중 몇 가지를 아래에 소개합니다.

① NRZ (Non-Return-to-Zero)

- 0은 0으로, 1은 1로 인코딩되는 방식입니다.
- 신호의 진폭이 변하지 않아 노이즈에 강합니다.
- 간단하고 구현이 쉬우며, 잡음에 대한 강건성이 높다.
- 시간에 따라 0과 1의 패턴이 반복되지 않아 클록 신호 추출이 어렵다.

② RZ (Return-to-Zero)

- 각 비트는 1/2의 동안 동일한 진폭을 가지고 있고, 나머지 1/2의 동안 0 진폭을 가지는 방식입니다.
- 신호의 각 비트가 구분될 수 있으므로 클록 복구가 용이합니다.
- NRZ보다는 구현이 복잡하며, 시간 구간이 두 배가 필요하다.

③ Manchester Encoding

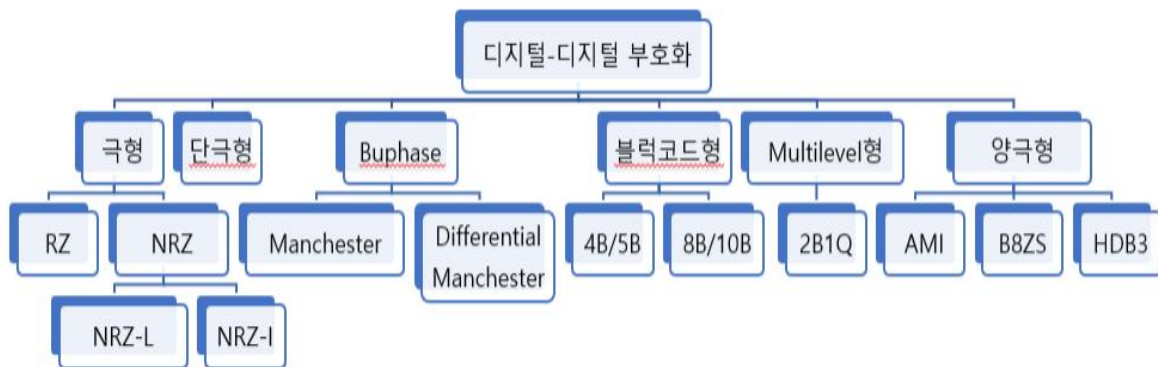
- 각 비트는 시간의 절반에 걸쳐서 1과 0을 번갈아가며 인코딩됩니다.
- 클록 복구가 용이하며 DC 바이어스 문제가 발생하지 않습니다.
- 클록 신호를 추출하는 데 필요한 동기화 비트를 제공한다.
- 디지털 신호가 전송되는 동안 변화가 지속되어 잡음이 발생하는 환경에서도 신호의 강건성이 높다.

④ AMI (Alternate Mark Inversion)

- 0에 대응하는 전압값은 0V이고, 1에 대응하는 전압값은 양수 전압값과 음수 전압값을 번갈아가며 사용하는 방식입니다. 예를 들어, 1에 대응하는 전압값을 5V와 -5V로 번갈아가며 사용한다면, 0은 계속 0V를 유지하고, 1은 5V와 -5V를 번갈아가며 사용합니다.

⑤ 4B/5B

- 4비트의 디지털 정보를 5비트로 인코딩하는 방식입니다.
- 잡음에 강한 장점이 있으며, 디지털 신호를 직접 변환하는 방식이 아닌 비트 패턴을 변환하는 방식이므로 다른 인코딩 방식과 결합하여 사용하기도 합니다.



[그림 3-33] 디지털-디지털 부호화 방식

위와 같은 방식 외에도 다양한 방식이 존재합니다. 선택할 방식은 전송 대역폭, 오류 처리 능력, 비용 등 다양한 요소를 고려하여 결정됩니다.

(나) 아날로그-디지털 부호화

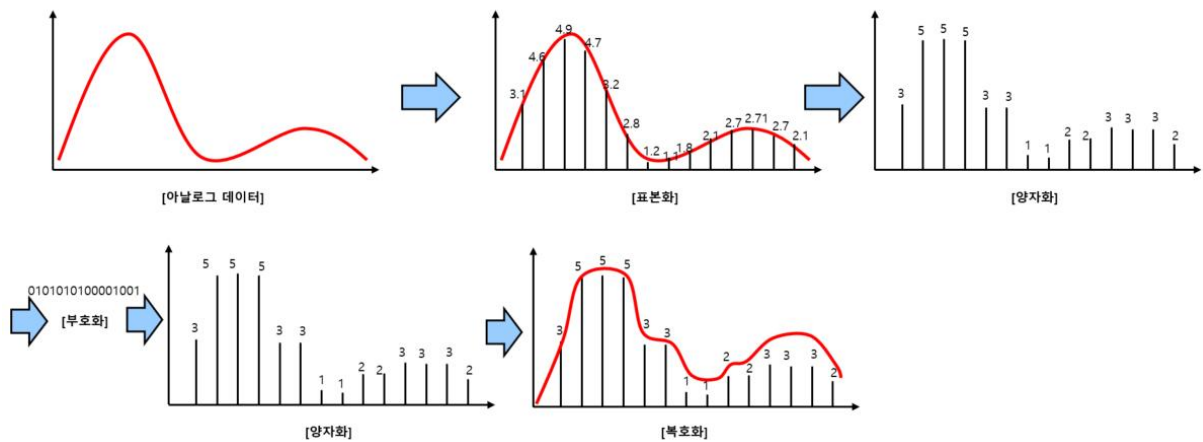
① Delta Modulation

Delta Modulation은 아날로그 신호의 기울기를 측정하여 디지털 신호로 변환하는 방식입니다. 샘플링 주기가 짧아 PCM보다 더 높은 전송 속도를 제공합니다. 그러나 잡음에 민감하며, 신호의 높은 주파수 구성 요소를 잃어버릴 수 있습니다.

② PCM (Pulse Code Modulation)

PCM은 아날로그 신호를 샘플링하여 디지털 신호로 변환하는 방식입니다. 샘플링은 아날로그 신호를 일정한 시간 간격으로 측정하여 디지털 값으로 변환하는 과정입니다. 이후 변환된 디지털 신호를 채널에 전송하고, 수신 측에서는 이를 다시 아날로그 신호로 복원하여 출력합니다. 이 방식은 대부분의 디지털 음성 통화에서 사용됩니다.

PCM에서는 아날로그 신호를 시간 축에 따라 일정한 간격으로 샘플링하여 이산적인 값으로 변환합니다. 이 샘플링된 값을 이진수로 표현하여 디지털 신호를 만듭니다. 이렇게 만들어진 디지털 신호는 이진수로 표현되기 때문에, 디지털 정보처리에 매우 유용하게 사용됩니다.



[그림 3-34] PCM(Pulse Code Modulation) 변조 과정

PCM (Pulse Code Modulation)은 아날로그 신호를 디지털 신호로 변환하는 기술 중 하나입니다. PCM 변조는 디지털 음성 통신, 오디오 녹음 및 음악 제작 등 많은 응용 분야에서 사용됩니다. PCM 변조는 아날로그 신호를 여러 개의 샘플로 분할하여 각 샘플의 크기를 디지털 코드로 변환합니다. 샘플링 레이트는 초당 샘플링 수를 나타내며, 높은 샘플링 레이트는 더 정확한 아날로그 신호의 표현을 가능하게 합니다.

PCM 변조 과정은 다음과 같습니다.

① 샘플링(Sampling)

PCM 변조에서 샘플링 과정은 아날로그 신호를 작은 구간으로 나누는 과정을 말합니다. 이 과정에서는 아날로그 신호를 일정한 시간 간격으로 샘플링하여 디지털 신호를 만들어냅니다. 이 시간 간격을 샘플링 주파수라고 합니다. 샘플링 주파수는 아날로그 신호의 주파수 범위에 따라 결정됩니다.

샘플링 주파수는 아날로그 신호의 최대 주파수보다 높아야 하며, 일반적으로는 최대 주파수의 2배 이상으로 설정됩니다. 샘플링 주파수가 너무 낮으면 아날로그 신호를 충분히 재현할 수 없으며, 너무 높으면 디지털 신호의 용량이 커지게 됩니다.

② 양자화(Quantization)

양자화는 아날로그 신호의 크기를 디지털 신호로 변환하는 과정에서 발생하는 신호 손실을 줄이기 위해 필요합니다. 이 과정에서는 아날로그 신호를 일정한 간격으로 나누어 그 구간에서의 아날로그 신호의 크기를 측정하고, 이 크기를 디지털 신호로 변환합니다.

이 때 변환된 디지털 신호는 일정한 값을 갖게 되는데, 이 값은 아날로그 신호의 크기를 일정한 간격으로 나누어 양자화한 결과물입니다. 양자화는 디지털 신호의 품질에 큰 영향을 미치기 때문에, 적절한 양자화 방식을 선택하는 것이 중요합니다. 일반적으로는 선형 양자화 방식을 많이 사용하며, 양자화 단계의 크기를 적절하게 조절하여 신호 품질과 양자화 오차를 균형있게 조절합니다.

③ 부호화(Encoding)

부호화는 디지털 신호를 처리하기 위해 필요한 과정 중 하나로, 디지털 신호를 이진 코드로 변환하는 것이 일반적입니다. PCM 변조에서는 양자화된 신호를 부호화하여 디지털 신호로 표현합니다.

PCM 변조는 신호를 정확하게 디지털화하지만, 데이터 비트의 수가 많아질수록 처리해야 할 데이터 양도 증가합니다. 이로 인해 실시간 전송이 필요한 많은 애플리케이션에서는 데이터 압축 알고리즘을 사용하여 처리량을 줄이는 것이 일반적입니다.

(다) 디지털-아날로그 부호화

브로드밴드(Broadband)란, 대역폭(Bandwidth)이 넓은 통신 기술을 의미합니다. 일반적으로, 1MHz 이상의 대역폭을 갖는 기술을 브로드밴드라고 합니다. 브로드밴드 기술은 주로 고속 인터넷, IPTV, 인터넷 전화 등에서 사용됩니다.

브로드밴드 기술은 아날로그 신호와 디지털 신호를 모두 사용할 수 있습니다. 아날로그 신호를 사용하는 브로드밴드 기술로는 케이블 모뎀(Cable Modem)이나 디지털 가입자 회선(Digital Subscriber Line, DSL)이 있습니다. 디지털 신호를 사용하는 브로드밴드 기술로는 광섬유 인터넷(Optical Fiber Internet)이나 무선 브로드밴드(Wireless Broadband)가 있습니다.

광섬유 인터넷은 광섬유 케이블을 사용하여 대역폭이 매우 넓은 인터넷 서비스를 제공합니다. 이 기술은 전송 속도가 빠르고 대역폭이 매우 넓으며, 장거리 통신에도 용이합니다. 하지만 설치 비용이 높은 단점이 있습니다.

무선 브로드밴드는 라우터와 같은 무선 장치를 사용하여 인터넷에 접속하는 기술입니다. 이 기술은 이동성이 높고 설치가 쉽다는 장점이 있습니다. 하지만 무선 신호의 범위가 제한되어 있고, 대역폭이 좁아서 전송 속도가 느릴 수 있습니다.

이러한 브로드밴드 기술은 대역폭이 넓어 빠른 인터넷 속도를 제공할 수 있습니다. 또한, 동시에 다수의 사용자가 인터넷에 접속하여 데이터를 이용할 수 있는 환경을 제공합니다. 하지만 이러한 브로드밴드 기술을 사용하기 위해서는 인프라가 충분히 갖추어져 있어야 하고, 이에 따른 초기 설치 비용과 유지 보수 비용이 높아질 수 있다는 단점도 있습니다.

디지털 정보를 아날로그 신호로 전송하는 경우에 해당되며, 정보를 아날로그 반송파 주파수로 변화하는 것을 보통 변조(Modulation)이라고 하며, 이를 다시 원래의 정보로 변환하는 것을 복조(Demodulation)이라고 한다.

디지털 정보를 아날로그 신호로 전송하는 방식에는 여러 가지가 있습니다. 여기서는 주로 사용되는 몇 가지 방식에 대해 설명하겠습니다.

① ASK (Amplitude-Shift Keying)

디지털 정보를 전송하기 위해 아날로그 신호의 진폭을 변경하여 전송하는 방식입니다. ASK는 저렴하고 단순한 구조로 빠른 속도를 제공하며, 대역폭을 절약할 수 있는 장점이 있습니다.

② FSK (Frequency-Shift Keying)

디지털 정보를 전송하기 위해 아날로그 신호의 주파수를 변경하여 전송하는 방식입니다. FSK는 모뎀 등에서 주로 사용되며, 속도가 느리고 대역폭이 넓은 단점이 있습니다.

③ PSK (Phase Shift Keying)

PSK는 직교한 캐리어 파형의 위상을 변경하여 디지털 신호를 아날로그 신호로 전송하는 방식입니다. 이 방식은 다른 방식보다 더 적은 대역폭을 사용하여 전송할 수 있습니다. 무선 통신에서 주로 사용됩니다.

④ QAM (Quadrature Amplitude Modulation)

QAM은 다중 채널을 사용하여 디지털 신호를 아날로그 신호로 전송하는 방식입니다. 이 방식은 다른 방식보다 더 높은 전송 속도를 제공하며, 대부분의 케이블 및 위성 통신에서 사용됩니다.

(라) 아날로그-아날로그 부호화

아날로그-아날로그 부호화(Analog-to-Analog Encoding)란, 아날로그 신호를 변조하는 방식이 아니라, 아날로그 신호 자체를 변화시켜 전송하는 방식을 말합니다. 이 방식은 아날로그 신호의 크기, 주기, 위상 등을 변화시켜 전송하는데 사용됩니다. 이 방식은 아날로그 신호를 디지털로 변환하는 것보다는 더 복잡하지만, 일부 응용 분야에서 여전히 유용합니다.

① AM (Amplitude Modulation):

AM 방식은 아날로그 신호를 캐리어 신호의 진폭을 변화시켜 전송하는 방식입니다. 이 방식은 아날로그 라디오 방송에서 가장 많이 사용됩니다. 변조 과정에서, 아날로그 신호의 진폭은 캐리어 신호의 진폭을 변화시키는데 사용됩니다. 변조된 신호는 송신기에서 전파 안테나를 통해 전파됩니다. 수신기는 캐리어 신호와 함께 변조된 신호를 수신하고, 변조된 신호에서 아날로그 신호를 복원합니다.

② FM (Frequency Modulation):

FM 방식은 아날로그 신호를 캐리어 신호의 주파수를 변화시켜 전송하는 방식입니다. 이 방식은 라디오 방송에서 주로 사용되며, AM 방식보다 음질이 더 좋고 잡음이 적은 장점이 있습니다. 변조 과정에서, 아날로그 신호의 크기는 변하지 않고, 신호의 주파수를 변조된 신호의 주파수로 매핑합니다. 변조된 신호는 송신기에서 전파 안테나를 통해 전파됩니다. 수신기는 캐리어 신호와 함께 변조된 신호를 수신하고, 변조된 신호에서 아날로그 신호를 복원합니다.

③ PM (Phase Modulation):

PM 방식은 아날로그 신호를 캐리어 신호의 위상을 변화시켜 전송하는 방식입니다. 이 방식은 무선 통신에서 사용되며, 주파수 변조보다는 더 복잡하지만 더 많은 정보를 전송할 수 있는 장점이 있습니다. 변조 과정에서, 아날로그 신호의 위상은 캐리어 신호의 위상을 변화시키는 데 사용됩니다. 변조된 신호는 송신기에서 전파 안테나를 통해 전파됩니다. 수신기는 캐리어 신호와 함께 변조된 신호를 수신하고, 변조된 신호에서 아날로그 신호를 복원합니다.

(1) 단말 장치: DTE(Data Terminal Equipment)

단말장치(DTE, Data Terminal Equipment)는 컴퓨터나 네트워크 장비와 같은 종단 장치를 의미합니다. DTE는 데이터를 생성하거나 소비하는 장치이며, 데이터 통신을 위한 물리적 또는 논리적 연결에 대한 책임을 지는 장치입니다. 예를 들어, 개인 컴퓨터나 노트북은 DTE로 분류됩니다. 이 장치들은 데이터를 생성하고 소비하며, 데이터 통신을 위한 물리적 또는 논리적 연결을 수립할 수 있습니다. 이와 반대로, 모뎀은 DCE(Data Communication Equipment)로 분류되며, DTE와 DCE 간의 통신을 가능하게 합니다.

DTE는 데이터 전송을 위한 다양한 프로토콜과 인터페이스를 지원합니다. 예를 들어, 시리얼 포트, USB, 이더넷 등 다양한 인터페이스를 사용하여 데이터를 전송할 수 있습니다. 이러한 인터페이스를 사용하여 DTE는 다른 DTE 또는 DCE와 통신하며, 데이터 전송 속도와 관련된 다양한 매개변수를 설정할 수 있습니다.

따라서, DTE는 컴퓨터 네트워크에서 중요한 역할을 합니다. 이를 통해 데이터가 생성되고 처리되며, 다른 장치와 통신할 수 있게 됩니다. 일반적으로 DTE에는 컴퓨터, 서버, 터미널, 프린터, 등 컴퓨터와 연결하는 모든 주변장치를 말한다.

(2) 신호 변환 장치: DCE(Data Circuit-terminating Equipment)

DCE(Data Circuit-terminating Equipment)는 데이터 통신 회선을 연결하는 장치를 말합니다. DCE는 데이터 통신 회선을 관리하고, DTE와 통신을 가능하게 합니다.

일반적으로, DCE는 네트워크의 중앙부에 위치하여, 여러 개의 DTE를 연결합니다. 이 때, DCE는 다양한 인터페이스를 사용하여 DTE와 통신할 수 있습니다. 예를 들어, 모뎀은 DCE의 한 종류로, 전화 회선을 사용하여 인터넷에 연결할 수 있도록 합니다. 또한, DCE는 데이터 통신 회선을 관리하는 역할을 합니다. 이를 통해 데이터 전송 속도, 전송 프로토콜, 오류 제어 등 다양한 매개변수를 설정할 수 있습니다. 이러한 설정은 데이터 전송의 안정성과 속도를 보장하는 역할을 합니다. 따라서, DCE는 데이터 통신 회선을 관리하고, DTE와의 통신을 가능하게 함으로써, 컴퓨터 네트워크에서 중요한 역할을 합니다. 이를 통해 데이터 전송이 안정적이고, 속도가 빠르게 이루어질 수 있습니다.

DCE(Data Circuit-terminating Equipment)는 데이터 회선 종단 장치로, 데이터 통신에 사용되는 장비로서 데이터의 전송, 수신, 제어 등의 역할을 수행합니다. 일반적으로 DCE는 다음과 같은 장비로 구성됩니다.

① 모뎀(Modem)

MOdulator와 DEModulator의 합성어로 전화회선, 케이블, 광케이블 등의 매체를 통해 디지털 신호를 모뎀에서 아날로그 신호로 변환하고, 아날로그 신호를 디지털 신호로 변환하여 다른 DCE 또는 DTE와 통신

② CSU/DSU(Channel Service Unit/Data Service Unit)

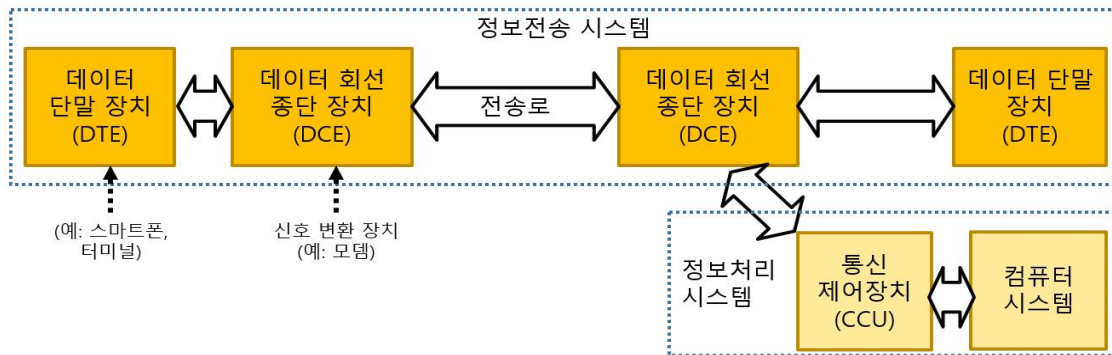
디지털 회선에서 디지털 신호를 전송하기 위해 필요한 신호 증폭 및 타이밍 조정 장비로서 T1, E1 등의 디지털 회선에서 사용

③ 라우터(Router)

다른 네트워크와 통신하기 위해 패킷을 전송하고 라우팅하는 데 사용되는 장비로 다른 DCE와 DTE 간의 데이터 흐름을 중개

④ 스위치(Switch)

여러 대의 컴퓨터 또는 네트워크 장비를 연결하여 데이터 통신을 가능하게 하는 장비



[그림 3-35] DTE/DCE의 관계

17.5. 정보교환 장치

정보 교환 장치는 데이터 통신에서 다양한 장치들 간에 정보를 교환하도록 돕는 장치를 일컫습니다. 이 장치는 데이터 통신에서 중요한 역할을 수행하며, 네트워크 기기들 간의 효율적인 통신을 가능하게 합니다. 다양한 종류의 정보 교환 장치가 있으며, 일부 예시는 다음과 같습니다:

① 스위치 (Switch)

LAN (Local Area Network) 상에서 데이터를 전송하는데 사용되며, 이더넷 프레임에 따라 목적지 장치로 데이터를 전송합니다.

② 라우터 (Router)

다양한 네트워크 간 데이터 전송을 가능하게 하는데 사용되며, IP 주소를 기반으로 데이터를 전송합니다.

③ 게이트웨이 (Gateway)

서로 다른 프로토콜로 작동하는 네트워크 간의 통신을 가능하게 합니다.

④ 모뎀 (Modem)

아날로그 신호를 디지털 신호로 변환하여 데이터 통신을 가능하게 합니다.

⑤ 허브 (Hub)

데이터를 전송하는데 사용되며, 송신되는 데이터를 연결된 모든 장치에 브로드캐스트합니다. 그러나 스위치와는 달리, 허브는 데이터 충돌 문제를 해결할 수 없습니다.

이외에도 다양한 정보 교환 장치들이 있으며, 각각의 기능과 용도에 따라 선택하여 사용됩니다.

18. 정보통신망

18.1. 정보통신망의 개념

일반적으로 데이터 통신에서 통신 단말들은 노드(Node)라고 부르며, 노드끼리 서로 연결하는 것을 링크(Link)라고 한다. 또한 링크를 통해 통신 경로의 끝과 끝이 연결되는 것을 경로(Path)라고 한다. 정보통신망이란 정보통신시스템에서 데이터/정보/지식을 송수신 노드 간에 약속된 수단 및 절차에 따라 정보를 주고받는 링크들로 구성되어 표현되는 하나의 시스템이다.

18.2. 정보통신망의 구성

정보통신망의 구성은 다음과 같은 형태로 분류할 수 있다.

[표 3-8] 정보통신망의 분류

| 분류 방식 | 설명 |
|-------------|-------------------------|
| 네트워크 범위와 연결 | LAN, MAN, WAN, VAN |
| 구성형태 | 성형, 버스형, 트리형, 망형 |
| 교환방식 | 회선 교환, 메시지 교환, 패킷 교환 |
| 전송대역과 속도 | 협대역망, 광대역망, 저속망, 고속망 |
| 서비스 제공 | 클라이언트/서버형, P2P형, 하이브리드형 |
| 전송 형식 | 아날로그망, 디지털망 |

정보통신망에서는 데이터가 전송되는 방식에 따라 다양한 교환 방식이 존재합니다. 다음은 정보통신망의 교환 방식에 따른 분류입니다.

① 회선 교환 방식 (Circuit Switching)

회선 교환 방식은 통화나 데이터를 전송하기 위해 두 노드 사이에 전용 회선을 연결하여 통신하는 방식입니다. 이 방식에서는 전용 회선을 연결하고 나면 해당 회선을 계속 사용하기 때문에 전송 속도가 일정합니다. 하지만 회선을 연결하면 해당 회선이 계속 점유되기 때문에 다른 통신 요청은 대기해야 하므로 효율성이 떨어지는 단점이 있으며 전화나 팩스 등에서 많이 사용됩니다.

② 패킷 교환 방식 (Packet Switching)

패킷 교환 방식은 데이터를 일정한 크기의 패킷으로 나누어 전송하는 방식입니다. 각 패킷은 독립적으로 전송되며, 전송 경로는 패킷마다 달라질 수 있습니다. 패킷 교환 방식은 회선 교환 방식과 달리 전송 경로가 유연하고 다수의 사용자가 동시에 통신할 수 있어 효율성이 높아 인터넷에서 많이 사용됩니다.

③ 메시지 교환 방식 (Message Switching)

메시지 교환 방식은 데이터를 전송하기 위해 메시지 단위로 나누어 전송하는 방식입니다. 메시지는 목적지까지 경유하는 중간 노드에서 저장하고 전달됩니다. 전송 시간은 패킷 교환 방식에 비해 더 길지만, 중간 노드에서 저장되기 때문에 전송 중간에 발생하는 오류나 중단 등의 문제를 해결할 수 있습니다. 이 방식은 오래전에 사용되었으나, 지금은 거의 사용되지 않습니다.

이러한 교환 방식은 통신망의 특성에 따라 적용되며, 이제는 대부분 패킷 교환 방식이 사용되고 있습니다.

또한, 네트워크 상의 웹 서비스 제공 형태에 따른 분류는 다음과 같다. 정보통신망에서는 다양한 모델을 사용하여 데이터 및 서비스를 분산 처리합니다. 여기에서는 클라이언트/서버 모델, P2P 모델, 하이브리드 모델에 대해 설명해 드리겠습니다.

① 클라이언트/서버 모델

클라이언트/서버 모델은 가장 일반적인 모델입니다. 이 모델은 클라이언트(사용자)와 서버(데이터/자원을 제공하는 기계) 간의 관계를 기반으로 합니다. 클라이언트는 서버에 요청을 보내고, 서버는 해당 요청을 받아 처리한 후 결과를 클라이언트에게 보내는 방식으로 작동합니다. 이 모델은 보안성과 안정성이 높고 중앙 집중화된 관리가 가능하다는 장점이 있으며 인터넷에서의 웹 서비스, 이메일, FTP 등에서 사용됩니다.

② P2P 모델

P2P 모델은 중앙 서버가 없이 모든 노드가 대등한 관계로 데이터와 서비스를 분산 처리하는 모델입니다. 이 모델에서는 각 노드가 서로 직접 통신하여 데이터를 주고 받습니다. 이 모델은 중앙 집중화된 관리가 없어 유연성이 높으며 서버 다운 등의 문제가 발생해도 전체 시스템에 큰 영향을 주지 않는다는 장점이 있습니다. P2P 모델은 클라이언트/서버 모델과 달리, 데이터/자원을 서로 공유하는 방식으로 작동합니다. 이 모델은 Peer-to-Peer 모델이라고도 불립니다. 각 클라이언트는 서로 독립적이며, 데이터/자원을 다른 클라이언트와 공유하고 다른 클라이언트로부터 데이터/자원을 받아들일 수 있습니다. 이 모델은 BitTorrent, Skype 등에서 사용됩니다.

③ 하이브리드 모델

하이브리드 모델은 클라이언트/서버 모델과 P2P 모델을 결합한 모델입니다. 이 모델에서는 서버가 중심적 역할을 하면서도 일부 노드가 직접 통신하여 데이터를 주고 받습니다. 이 모델은 클라이언트/서버 모델의 안정성과 보안성을 유지하면서 P2P 모델의 유연성과 분산 처리의 장점을 결합한 모델입니다. 이 모델은 대규모 파일 공유 등에 많이 사용됩니다. 예를 들어, Dropbox는 클라이언트/서버 모델을 사용하면서 동시에 P2P 기술을 활용해 파일 전송 속도를 향상시키고 있습니다.

18.3. 선박통신

선박통신은 선박 내부 및 선박 간 통신을 의미합니다. 선박통신 시스템은 선박 안전 및 운영을 지원하기 위해 설계되었으며, 다양한 통신 수단을 통해 선박 간 정보를 교환할 수 있습니다.

선박통신 시스템에는 크게 무선 통신과 유선 통신 두 가지 종류가 있습니다. 무선 통신은 선박과 해안 간, 선박과 선박 간, 그리고 위성을 통한 통신 등 다양한 방법으로 이루어집니다. 대표적으로 VHF, HF, UHF, 휴대용 무선기 등이 있습니다. 이러한 무선 통신 시스템은 선박 운항 중 정보를 교환하는 데 필수적입니다. 예를 들어, 선박은 AIS(자동식별체계)를 통해 근처 다른 선박의 위치와 속도 정보를 수신하고, GMDSS(글로벌 해양 안전통신체계)를 통해 긴급 상황 발생 시 도움을 요청하거나 정보를 전달합니다.

유선 통신은 주로 선박 내부에서 사용되며, 전화, 인터넷, CCTV 등 다양한 용도로 활용됩니다. 선박 내부에서는 일반적으로 LAN(Local Area Network)이 구축되어 있으며, 다양한 장비들이 LAN을 통해 연결됩니다. 또한, 인터넷은 선박 내부의 크루 및 승객들이 정보를 검색하거나 연락을 주고 받는 데에도 사용됩니다.

선박통신은 안전한 항해 및 선박 운영을 위해 국제해사기구(IMO)와 국제전기통신연합(ITU) 등 국제기구에서 규제되고 있습니다. 예를 들어, GMDSS(Global Maritime Distress and Safety System)는 선박 간의 긴급상황에서 사용되는 긴급 구조 체계로, 전 세계적으로 표준화되어 있습니다. 또한, 선박통신은 선박 운영을 지원하는 다양한 정보를 전송하는 역할을 합니다. 이러한 정보는 항로, 날씨, 파도, 조류, 해수면 온도 등의 정보와 선박 운영에 필요한 경제적, 법적, 기술적 정보 등이 포함됩니다.

마지막으로, 선박통신은 국제무역 및 물류 산업에서 중요한 역할을 합니다. 선박통신은 선박 운영자들이 물류와 국제무역에 필요한 정보를 교환할 수 있도록 해주며, 이는 국제무역 및 물류 산업의 효율성을 높이는 데 큰 역할을 합니다.

18.4. 해상통신

해상통신은 해양 환경에서 정보 및 데이터를 교환하는 모든 형태의 통신을 말합니다. 이는 선박 간, 선박과 육지 간, 선박과 위성 간, 그리고 해양 기반 시설물과 육지 간의 통신 등을 포함합니다. 해상통신은 선박 운영, 해양 자원 개발, 해양 안전, 해양 보호, 기상 및 해양 정보 등 다양한 분야에서 중요한 역할을 합니다. 해상에서 선박이 운용되는 동안에는 다양한 통신 수단이 사용됩니다. 이러한 통신 수단은 선박 운항에 필수적이며, 선박 운영에 대한 다양한 정보를 전달하고, 안전한 항해 및 응급 상황에서의 대처를 위한 수단으로 사용됩니다.

가장 일반적으로 사용되는 해상통신에는 다음과 같은 것들이 있습니다.

① VHF(Very High Frequency) 무선

VHF 무선은 156~174MHz의 주파수 대역을 사용하며, 선박 간의 통신 및 해안국과의 통신에 사용됩니다. 주로 운항 상황 보고, 조난신고, 날씨 정보, 항해 정보 등을 전달하는 용도로 사용됩니다.

② MF(High Frequency) / HF(High Frequency) 무선

MF/HF 무선은 2~30MHz의 주파수 대역을 사용하여 선박과 지상국, 다른 선박과의 통신에 사용됩니다. MF/HF 무선은 VHF 무선과는 달리 장거리 통신이 가능하며 대기권 상태에 따라 멀리 떨어진 지역과도 통신이 가능합니다.

③ 위성 통신

위성 통신은 선박에서 긴거리 통신이 필요한 경우에 사용됩니다. 인터넷, 전화, 이메일 등 다양한 정보 교환을 위해 사용됩니다.

④ AIS(Automatic Identification System)

AIS는 선박 간의 위치, 속도, 방향 등의 정보를 자동으로 교환하는 시스템으로, 안전한 항해와 선박 운영에 매우 중요한 역할을 합니다.

⑤ GMDSS(Global Maritime Distress and Safety System)

GMDSS는 국제 연안수안공간조약(SOLAS)에 따라 응급 상황에서 선박이 발신하는 신호를 전파로 지상국에 전달하는 시스템입니다. 이 시스템은 조난상황, 해저 지진 등의 긴급 상황에서 선박 승무원의 생명과 안전을 보호하기 위해 국제적으로 규제되고 있습니다.

⑥ 네비게이션 시스템

GPS(Global Positioning System)와 GLONASS(Global Navigation Satellite System) 등의 위성 네비게이션 시스템은 선박의 위치 정보를 파악하는 데 사용됩니다.

⑦ 라디오 방향탐지기(RDF)

RDF는 무선 신호를 이용하여 선박의 위치를 파악하는데 사용됩니다.

⑧ VTS (Vessel Traffic Service)

선박의 운항 상황과 위치, 해양 환경 등을 모니터링하고, 선박 간의 충돌을 예방하며, 안전한 항해를 돕기 위해 다양한 정보를 제공합니다. 이러한 정보는 선박 운항 관리 시스템, AIS, 레이더, CCTV 등 다양한 항만통신 시스템을 통해 수집되며, VTS 센터에서 관제됩니다.

18.5. 항만통신

항만통신은 선박 운항과 관련된 다양한 정보를 교환하고, 안전한 항만 운영을 위한 통신 시스템입니다. 선박이 항만에 들어왔을 때 사용되는 항만통신 시스템에는 다음과 같은 것들이 포함됩니다:

① 선박-항만 교신 시스템(VHF)

선박과 항만 간에 음성으로 통신할 수 있는 시스템입니다. 선박과 항만 간에 통상적으로 사용되는 주파수는 156~162MHz 사이입니다. VHF 시스템을 통해 항만 운영자는 선박의 도착 예정 시간과 위치, 무엇보다도 선박과의 안전 거리를 유지하기 위한 정보 등을 확인할 수 있습니다.

② AIS (Automatic Identification System)

AIS는 자동식별시스템으로, 선박의 위치, 속도, 방향 등의 정보를 교환하는 시스템입니다. AIS 시스템을 통해 항만 운영자는 선박의 운항 상황을 실시간으로 파악할 수 있습니다.

③ 레이더(Radar)

레이더는 선박의 위치와 움직임을 탐지할 수 있는 장비입니다. 항만 운영자는 레이더를 이용해 선박의 위치와 상황을 파악할 수 있습니다.

④ 항만 정보 시스템(Port Information System)

항만 정보 시스템은 선박 운항과 관련된 정보를 종합적으로 관리하는 시스템입니다. 선박의 도착 예정 시간, 화물 정보, 차세대항만계획 등의 정보를 제공합니다.

⑤ CCTV (Closed Circuit Television)

항구내의 시설물 및 선박의 상황을 모니터링하는 시스템입니다. 항만내의 안전을 유지하기 위해 사용됩니다.

⑥ VTS (Vessel Traffic Service)

선박의 운항 상황과 위치, 해양 환경 등을 모니터링하고, 선박 간의 충돌을 예방하며, 안전한 항해를 돕기 위해 다양한 정보를 제공합니다. 이러한 정보는 선박 운항 관리 시스템, AIS, 레이더, CCTV 등 다양한 항만통신 시스템을 통해 수집되며, VTS 센터에서 관제됩니다.

Ⅷ. 네트워크 기본

목포해양대학교 **홍성화** 교수



19. 네트워크 기초

19.1. 프로토콜의 개념

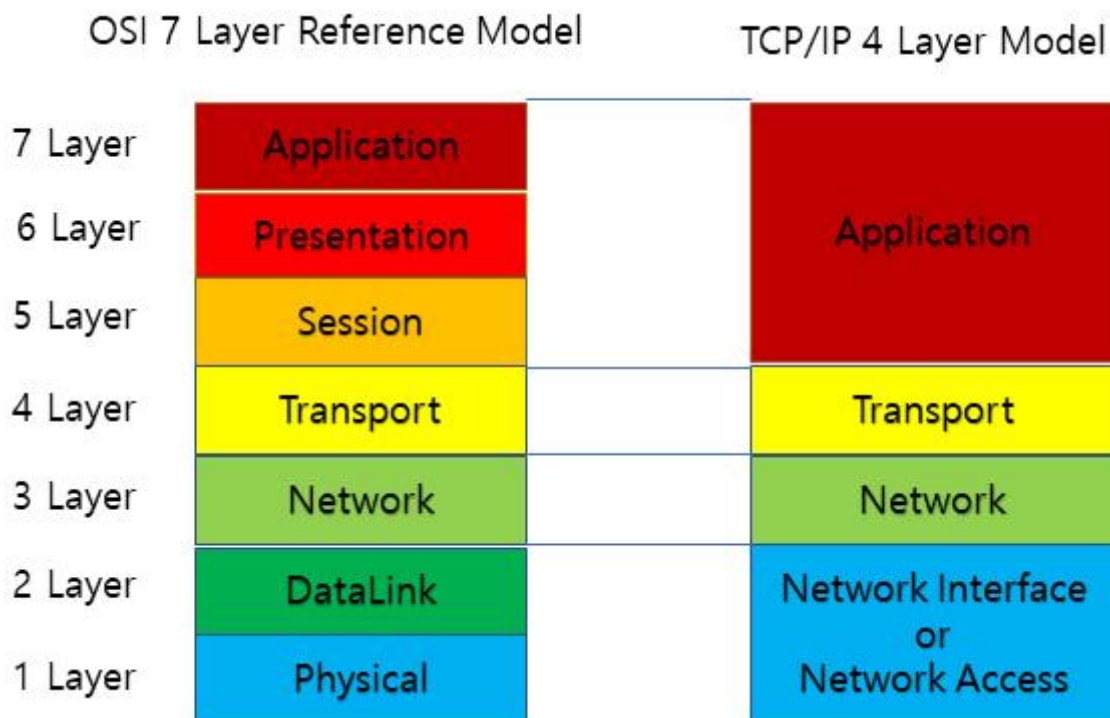
컴퓨터 네트워크에서 프로토콜(Protocol)은 컴퓨터 간의 통신에 사용되는 규약이나 규칙의 모임을 의미하며, 이 규칙은 송신자와 수신자 간에 서로 이해할 수 있는 언어로 작성되어 있으며, 정보를 안정적이고 안전하게 전송하는 데 필요한 모든 요소를 정의합니다. 즉, 프로토콜은 컴퓨터 간의 데이터 통신을 원활하게 하기 위한 통신 규약이라고 할 수 있습니다. 프로토콜은 데이터를 전송할 때 사용되는 통신 방법, 데이터 전송 속도, 데이터 압축 방식 등 다양한 기능을 포함하고 있습니다. 또한, 프로토콜은 컴퓨터 간의 통신에 사용되는 프로그램의 일부로 구현되어 있습니다.

프로토콜은 보통 7개의 계층으로 구성된 OSI(Open Systems Interconnection) 모델에 따라 설계됩니다. 이 모델은 각 계층이 서로 독립적이고, 서로 다른 기능을 수행하며, 서로 다른 프로토콜을 사용할 수 있다는 가정에 기반합니다. OSI 모델의 각 계층은 다음과 같습니다.

- 물리 계층(Physical Layer): 데이터 전송에 사용되는 물리적 장치와 전기 신호, 무선 주파수 등을 다룸
- 데이터 링크 계층(Data Link Layer): 전송된 데이터의 흐름을 제어하고 오류 검출 및 수정을 수행
- 네트워크 계층(Network Layer): 다른 네트워크 간의 데이터 전송 및 경로 선택을 담당합니다.
- 전송 계층(Transport Layer): 애플리케이션 간의 연결을 제공하고, 데이터 전송의 신뢰성을 유지
- 세션 계층(Session Layer): 컴퓨터 간의 대화 관리 및 동기화를 담당
- 표현 계층(Presentation Layer): 데이터를 이해하기 쉽도록 형식을 변환하고, 암호화 및 복호화를 수행
- 응용 계층(Application Layer): 사용자가 실제로 사용하는 서비스(이메일, 파일 전송 등)를 제공

프로토콜은 각 계층에서 사용되며, 계층 간에는 특정 규칙에 따라 데이터를 전달합니다. 예를 들어, TCP/IP 프로토콜은 인터넷에서 가장 일반적으로 사용되는 프로토콜이며, 전송 계층에서 신뢰성 있는 데이터 전송을 제공합니다. HTTP(HyperText Transfer Protocol)는 응용 계층에서 사용되는 프로토콜로, 웹 브라우저와 웹 서버 간의 통신을 관리합니다.

19.2. OSI-7 계층



[그림 3-36] OSI 7 Layer & TCP/IP 4 Layer

1계층 - 물리계층(Physical Layer)

- 주로 전기적, 기계적, 기능적인 특성을 이용해서 통신 케이블로 데이터를 전송
- 사용되는 통신 단위는 비트이며 이것은 1과 0으로 표시
- 단지 데이터를 전달만 할뿐 전송하려는(또는 받으려는)데이터가 무엇인지, 어떤 에러가 있는지 등에는 전혀 신경 쓰지 않음
- 단지 데이터 전기적인 신호로 변환해서 주고받는 기능
- 대표적인 장비는 통신 케이블, 리피터, 허브등

2계층 - 데이터 링크계층(DataLink Layer)

물리계층을 통해 송수신되는 정보의 오류와 흐름을 관리하여 안전한 정보의 전달을 수행할 수 있도록 도와주는 역할을 한다. 따라서 통신에서의 오류도 찾아주고 재전송도 하는 기능을 가지고 있는 것이다. 이 계층에서는 MAC 주소를 가지고 통신하게 되며, 이 계층에서 전송되는 단위를 프레임이라고 하고, 대표적인 장비로는 브리지, 스위치 등이 있다.(여기서 MAC주소를 사용한다.)

3계층 - 네트워크 계층(Network Layer)

이 계층에서 가장 중요한 기능은 데이터를 목적지까지 가장 안전하고 빠르게 전달하는 기능(라우팅)이다. 여기에 사용되는 프로토콜의 종류도 다양하고, 라우팅하는 기술도 다양하다. 이 계층은 경로를 선택하고 논리 주소(IP 주소)를 정하고 경로에 따라 패킷을 전달해주는 것이 이 계층의 역할이다. 이 계층의 대표적인 장비는 라우터이며, 요즘은 2계층의 장비 중 스위치라는 장비에 라우팅 기능을 장착한 Layer 3 스위치도 있다.

4계층 - 전송 계층(Transport Layer)

통신을 활성화하기 위한 계층이다. 보통 TCP프로토콜을 이용하며, 포트를 열어서 응용프로그램들이 전송을 할 수 있게 한다. 만약 데이터가 왔다면 4계층에서 해당 데이터를 하나로 합쳐서 5계층에 던져 준다. 단대단 오류제어 및 흐름제어 이 계층 까지는 물리적인 계층에 속한다.(TCP/UDP프로토콜을 사용한다.) 전송 계층(Transport layer)은 양 끝단(End to end)의 사용자들이 신뢰성있는 데이터를 주고 받을 수 있도록 해 주어, 상위 계층들이 데이터 전달의 유효성이나 효율성을 생각하지 않도록 해준다. 시퀀스 넘버 기반의 오류 제어 방식을 사용한다. 전송 계층은 특정 연결의 유효성을 제어하고, 일부 프로토콜은 상태 개념이 있고(stateful), 연결 기반(connection oriented)이다.

5계층 -세션 계층(Session Layer)

데이터가 통신하기 위한 논리적인 연결을 말한다. 통신을 하기위한 대문이라고 보면 된다.

하지만 4계층에서도 연결을 맺고 종료할 수 있기 때문에 우리가 어느 계층에서 통신이 끊어 졌나 판단하기는 한계가 있다. 그러므로 세션 계층은 4 계층과 무관하게 응용 프로그램 관점에서 봐야 한다.

세션 설정, 유지, 종료, 전송 중단시 복구 등의 기능이 있다. 세션 계층(Session layer)은 양 끝단의 응용 프로세스가 통신을 관리하기 위한 방법을 제공한다. 동시 송수신 방식(duplex), 반이중 방식(half-duplex), 전이중 방식(Full Duplex)의 통신과 함께, 체크 포인팅과 유휴, 종료, 다시 시작 과정 등을 수행한다. 이 계층은 TCP/IP 세션을 만들고 없애는 책임을 진다.

6계층 - 표현 계층(Presentation Layer)

데이터 표현이 상이한 응용 프로세스의 독립성을 제공하고, 암호화 한다. 표현 계층(Presentation layer)은 코드 간의 번역을 담당하여 사용자 시스템에서 데이터의 형식상 차이를 다루는 부담을 응용 계층으로부터 덜어 준다. MIME 인코딩이나 암호화 등의 동작이 이 계층에서 이루어진다.

7계층 - 응용 계층(Application Layer)

최종 목적지로서 HTTP, FTP, SMTP, POP3, IMAP, Telnet 등과 같은 프로토콜이 있다. 해당 통신 패킷들은 방금 나열한 프로토콜에 의해 모두 처리되며 우리가 사용하는 브라우저나, 메일 프로그램은 프로토콜을 보다 쉽게 사용하게 해주는 응용프로그램이다.

19.3. TCP/IP

TCP/IP는 인터넷 프로토콜 스위트(Internet Protocol Suite)의 핵심이 되는 프로토콜이다. TCP/IP는 데이터 통신을 위한 프로토콜로, 네트워크 계층과 전송 계층의 두 가지 계층으로 구성되어 있다.

TCP/IP는 데이터의 전송과 수신에 있어서 기능적으로 두 가지 프로토콜인 TCP (Transmission Control Protocol)와 IP (Internet Protocol)를 사용한다. IP는 패킷 전달과 라우팅을 담당하며, TCP는 데이터를 패킷으로 나누어 전송하고, 패킷 손실이나 중복 전송을 감지하여 문제를 해결한다.

TCP/IP는 인터넷에서 데이터의 안정적인 전송을 보장하는 가장 중요한 프로토콜로, 웹 브라우저나 이메일 클라이언트 등의 다양한 애플리케이션에서 사용됩니다. 또한, TCP/IP는 다른 네트워크 프로토콜과 상호 운용성을 제공하여, 다양한 네트워크 환경에서 데이터의 전송을 가능하게 합니다.

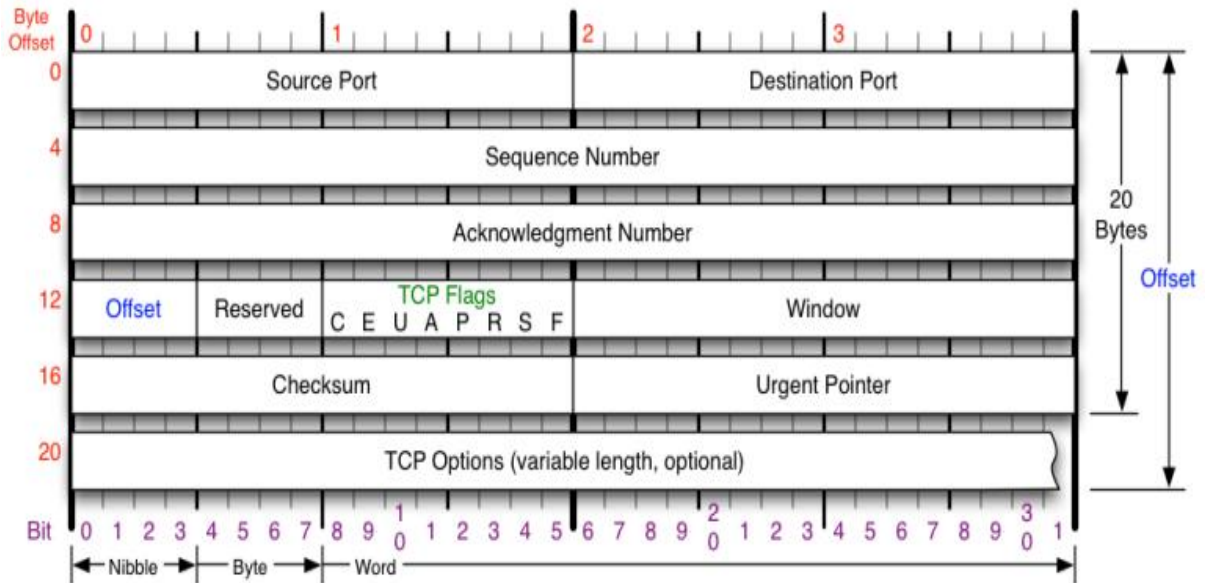
(1) TCP

TCP는 Transmission Control Protocol의 약자로, 인터넷 프로토콜 스위트(IP Suite)의 하나로서, 인터넷 상에서 데이터를 신뢰성 있게 전송하기 위해 사용되는 프로토콜입니다. TCP는 연결 지향형 프로토콜로, 세그먼트 단위로 데이터를 나누어 전송하며, 전송 과정에서 데이터 손실이나 손상이 발생하면 해당 세그먼트를 다시 전송합니다. 이를 통해 데이터 전송의 신뢰성을 보장합니다.

윈도우를 이용한 흐름제어와 ACK를 이용한 에러제어를 한다. 장점은 세그먼트의 전달이 보장된다는 것이며, 단점은 연결을 위한 초기 설정 시간이 필요하다는 것이다.

TCP의 PDU는 Segment로 애플리케이션의 메시지를 순차적으로 분할하고 데이터가 전후 상관 관계가 고려되어 있다는 의미에서 Segment 라는 용어를 사용한다고 한다.

TCP는 3-way handshake 방식을 사용하여 연결을 설정하며, 데이터 전송이 완료되면 4-way handshake 방식으로 연결을 종료합니다. 이 과정에서 TCP는 상대방과의 연결 상태를 확인하고, 패킷의 손실을 검출하고, 재전송 및 순서 제어를 수행합니다.



[그림 3-37] TCP Header

TCP 헤더는 TCP 패킷의 제어 정보를 담고 있으며, 총 20바이트로 구성됩니다. TCP 헤더는 다음과 같은 필드로 구성됩니다.

- Source port(발신지 포트 주소)
 - 출발지 포트번호
- Destination port(목적지 포트 주소)
 - 목적지 포트번호
- Sequence Number
 - 데이터가 잘 전송되는지를 보장하기 위해 들어가는 필드로, 전송되는 세그먼트의 순서번호

- Acknowledgment Number

- 통신상대가 발송한 패킷을 받았다는 것을 상대방에게 확인해주는 번호로 수신한 마지막 byte 순서번호에서 1을 더함

- Header Length

- TCP 헤더의 크기를 나타내는 것으로 4byte단위이다.

- Flags

- URG: Urgent Pointer가 유효함을 표시
- ACK: Acknowledgment Number가 유효함을 표시
- PSH: 수신자는 패킷데이터를 최대한 응용프로그램에서 전달할 것을 지시함(버퍼에 모아두지말고 다 보내라는 뜻)
- RST: 연결을 리셋하도록 지시
- SYN: 연결의 시작을 나타내기 위해 사용
- FIN: 연결을 종료하도록 지시

- Window Size

- 상대방 장비에게 얼마만큼의 데이터를 보내도 되는지를 알려주는 필드

- Checksum

- TCP 헤더와 데이터 필드에 대한 계산된 체크섬으로 에러 여부를 체크함

- Urgent Pointer

- URG flag가 On되어 있을 경우에만 유효(전송 데이터 중 긴급한 데이터의 위치 선정)

- Option

- 현재 정의되어 있는 옵션으로 최대 40byte까지 추가할 수 있다.

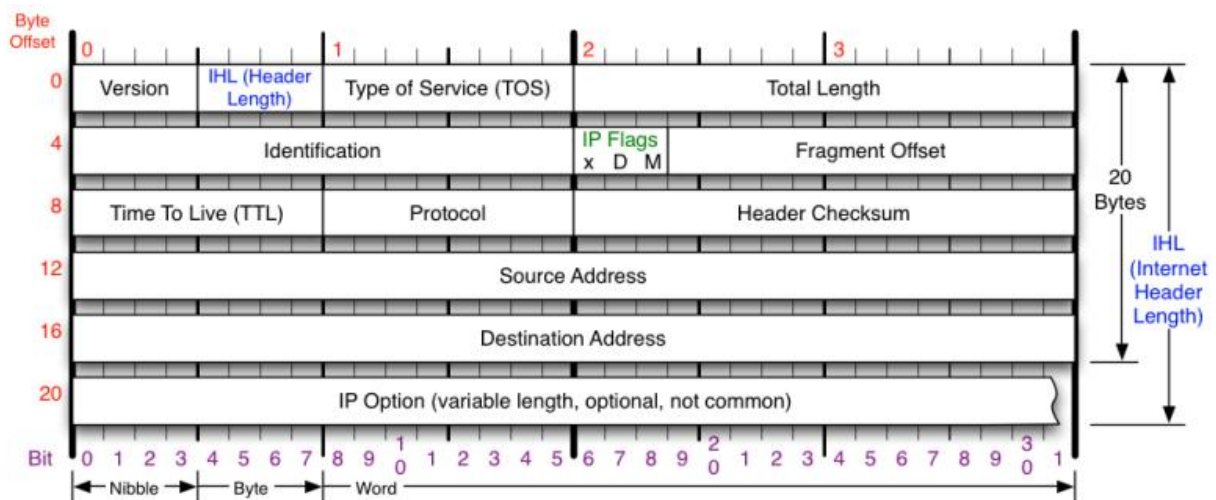
(2) IP

인터넷 프로토콜(Internet Protocol, IP)은 인터넷에서 컴퓨터나 기타 장치들이 데이터를 주고 받기 위한 통신 규약 중 하나입니다. IP는 인터넷 상에서 데이터를 전송하는데 있어서 패킷이라는 단위로 데이터를 나누어 전송하며, 이를 통해 데이터 전송을 효율적으로 처리할 수 있습니다.

IP는 OSI 7계층 모델에서 네트워크 계층에서 동작하며, 인터넷 상에서 패킷을 전송하고 라우팅하는 역할을 담당합니다. IP는 전송 계층에서 사용되는 TCP(Transmission Control Protocol)나 UDP(User Datagram Protocol)와 함께 사용되어 데이터의 안정성과 신속한 전송을 보장합니다.

IP는 인터넷에서 데이터를 분할하고 패킷으로 전송하는 역할을 합니다. 패킷은 데이터를 전송하기 위한 최소한의 단위로, 목적지 주소, 출발지 주소, 데이터, 그리고 제어 정보를 포함합니다. 이러한 패킷은 다른 네트워크에 전송될 때에도 사용됩니다.

IP는 또한 인터넷 주소 체계를 정의하고 있습니다. 이러한 주소는 "IP 주소"라고 하며, IPv4와 IPv6 두 가지 주소 체계가 있습니다. IPv4는 32비트로 구성된 주소이며, 예를 들어 "192.168.0.1"과 같이 숫자로 이루어져 있습니다. 반면, IPv6는 128비트로 구성된 주소이며, 예를 들어, "2001:0db8:85a3:0000:0000:8a2e:0370:7334"과 같이 숫자와 문자로 이루어져 있습니다. IP는 인터넷에서 중요한 역할을 담당하고 있으며, 인터넷의 기본적인 동작 원리를 이해하는 데 필수적인 개념 중 하나입니다.



[그림 3-38] IP Header

- 버전 : IP 프로토콜 버전. 현재는 4를 사용하고 있음
- IHL(헤더 길이) : IP 헤더의 길이
- ToS(서비스 유형) : 패킷을 운반할 때 우선순위 등을 나타냄
 - 처음에는 사양이 모호하여 최근엔 DiffServ라는 사양으로 필드 사용법을 재정의함
- 전체 길이 : IP 메시지 전체 길이를 나타냄
- ID 정보(identification) : 패킷을 식별하는 번호로 보통 패킷의 참조 번호가 여기에 기록됨
- IP 클라이언트에 의해 분할된 패킷(Fragmentation 된 패킷)은 모두 ID 정보 값이 같음
- 플래그 : 1비트로 조각나누기(Fragmentation)가 가능한지 나타내고 1비트로 현재 패킷이 나눠

진 패킷인지 나타냄

필드는 3비트이지만 2비트만 사용함

DF : 1일 경우 Fragmentation이 불가능한 패킷

MF : Fragmentation 패킷 중 마지막 패킷에 0으로 기록하여 더 이상 합칠 Fragmentation 패킷이 없음을 알림

- 프래그먼트 오프셋(Fragment Offset) : Fragmentation 되기 전 데이터 시작 점으로부터의 위치
- 생존 기간(TTL) : 라우터를 경유할 때마다 이 값이 1씩 감소하여 0이 되면 패킷이 폐기됨
네트워크 루프 때문에 패킷이 영구적으로 순환되는 것을 막기 위한 용도
- 프로토콜 번호
 - TCP : 06
 - UDP : 11
 - ICMP : 01
- 헤더 체크섬 : 오류 검사용 데이터지만 현재는 사용하지 않음
- 송신처 IP 주소 : 패킷 송신처의 IP 주소
- 수신처 IP 주소 : 패킷 수신처의 IP 주소
- 옵션 : 위에 설명하지 않은 제어 정보를 담기 위한 공간이지만 거의 사용하지 않음

19.4. IP 주소

IP주소(Internet Protocol address)는 인터넷상에 있는 컴퓨터의 고유한 주소로 이것을 바탕으로 인터넷 상의 한 컴퓨터에서 다른 컴퓨터로 데이터를 주고 받을 수 있다. 현재 사용되는(위에 보이는) IP 주소는 버전4(IPv4)로 점으로 나누어진 최대 12자리의 숫자로 이루어져 있다.

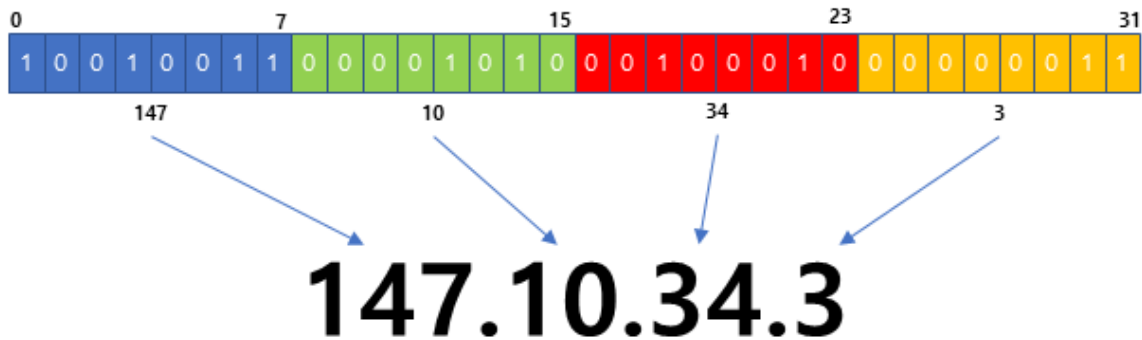
IP 주소(IP address, 표준어: 인터넷규약주소)는 컴퓨터 네트워크에서 장치들이 서로를 인식하고 인터넷에 연결된 모든 호스트들을 중에 통신할 상대를 식별하기 위한 고유주소로 OSI 7계층 중에서 3계층인 네트워크 계층에서 생성 및 사용된다.

현재 IP(IPv4)는 전송 측과 목적지 주소 영역이 각각 32비트로(약 43억 개) 구성되고, IPv6인 경우 128비트로(약 1조 개) 확장한다. 이진수 32비트로 이루어진 IP주소는 8비트 4개의 영역으로 구분하여 10진수로 표현한다

IP 주소를 줄여서 IP라고 부르기도 하나 IP는 인터넷 규약 자체를 가리키는 용어이므로 엄밀하게는 구별해야 한다. IP와 IP 주소는 다른 개념이다.

(1) IPv4

IPv4 주소는 전화번호와 같이 국내에서 표준을 정하고 정책을 수립하여 이용자에게 무한히 할당할 수 있는 자원이 아니라 전 세계적으로 관리되는 유한한 자원이다(약 43억개). 일부는 특수한 목적으로 예약되었으며, 주소 규정에 의하여 사용이 제한적이기 때문에 IP주소 할당 정책에 따라 부여하여 사용합니다.



[그림 3-34] IPv4 주소

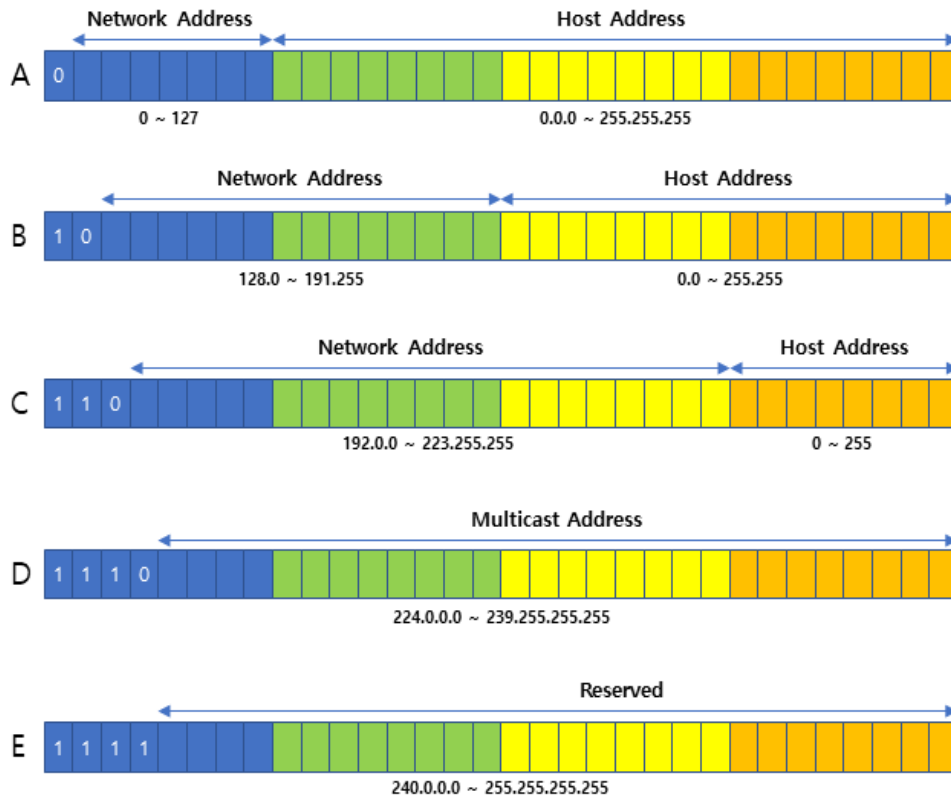
IPv4는 인터넷 프로토콜의 버전 4로, 인터넷에서 데이터를 전송하기 위한 주요 프로토콜 중 하나이다. IPv4는 32비트 길이의 IP 주소 체계를 사용하며, 이를 통해 인터넷 상에서 각 컴퓨터나 장치들을 식별한다.

IPv4의 IP 주소는 4개의 8비트 숫자(바이트)로 구성되어 있으며, 각 바이트는 점으로 구분되어 표시한다. IPv4는 인터넷에서 데이터를 전송하기 위한 핵심 프로토콜 중 하나이며, 다른 프로토콜인 TCP, UDP와 함께 사용된다.

TCP는 데이터 전송의 신뢰성을 보장하고, UDP는 데이터 전송의 신뢰성을 보장하지 않지만 빠른 전송이 필요한 경우에 사용된다. 됩니다.

IPv4는 32비트 IP 주소 체계를 사용하여 네트워크 상에서 패킷의 출발지와 목적지를 식별합니다. IPv4 주소는 네트워크 ID와 호스트 ID로 구성됩니다. 32비트 중에서 첫 번째 몇 개의 비트가 네트워크 ID를 결정하고, 나머지 비트가 호스트 ID를 결정합니다.

이러한 방식으로, IPv4는 다양한 네트워크에서 호스트들을 구분할 수 있으며, A, B, C 클래스는 일반 사용자에게 부여하는 네트워크 구성용, D 클래스는 멀티캐스트용, E 클래스는 향후 사용을 위하여 예약된 주소입니다.



[그림 3-35] IPv4 주소 클래스

IPv4는 인터넷에서 가장 널리 사용되는 프로토콜 중 하나이지만, 일부 한계점을 가지고 있습니다. 첫째, IPv4 주소의 32비트는 현재 인터넷에서 사용되는 장치 수를 제한하고 있습니다. 둘째, IPv4 주소는 기본적으로 보안 기능이 없으며, 패킷에 대한 무결성을 보장하지 않습니다.

이러한 한계점을 극복하기 위해 IPv6라는 새로운 버전의 인터넷 프로토콜이 개발되었습니다. IPv6는 128비트 IP 주소를 사용하여 더 많은 장치를 지원하고, 패킷의 무결성과 인증을 보장하는 기능을 제공합니다. 그러나 IPv6가 보급되기 전까지, IPv4는 여전히 많은 네트워크에서 사용되고 있습니다.

IPv4는 이제 구식으로 여겨지며, 다른 IP 버전인 IPv6로 대체되고 있습니다. IPv6는 더 많은 비트를 사용하여 IP 주소 공간을 확장하고, 보안 및 기능 면에서 더 많은 기능을 제공합니다. IPv4와 IPv6는 상호 운용이 가능하므로 현재 인터넷에서는 두 버전 모두 사용됩니다.

(2) IPv6

IPv6는 인터넷 프로토콜의 최신 버전으로, IPv4에서 겪었던 주소 고갈 문제와 보안 문제 등을 극복하기 위해 개발된 프로토콜입니다. IPv6는 128비트 IP 주소 체계를 사용하여 네트워크 상에서 패킷의 출발지와 목적지를 식별합니다.

IPv6 주소는 IPv4와는 다르게 16비트씩 8개의 필드로 나누어져 있습니다. 이러한 구조로 인해 IPv6는 무수히 많은 IP 주소를 제공합니다. IPv6의 주소 공간은 2의 128승, 즉 약 3400억억억억개의 IP 주소를 제공합니다.

따라서 IPv6는 인터넷 확장과 이동 기기 증가에 대응하기 위한 필수적인 기술입니다.

또한, IPv6는 데이터의 보안성과 무결성을 강화하기 위해 IPsec 기능을 내장하고 있습니다. 이로 인해 IPv6 기반의 통신에서는 데이터의 암호화와 인증 등이 가능해지며, 더욱 안전한 인터넷 환경을 제공할 수 있습니다.

IPv6는 또한 이동성 기능도 강화되어 있습니다. 이동성은 모바일 기기에서 자주 발생하는 문제 중 하나로, 이를 해결하기 위해 IPv6에서는 기기가 이동해도 주소가 변경되지 않도록 하는 "확장 가능한 주소 지정" 기능을 제공합니다.

마지막으로 IPv6는 다양한 네트워크 장비들과 호환성이 좋습니다. IPv6는 다양한 장비에서 지원되며, IPv4와 호환성을 유지하면서 IPv6로 전환하는 기술도 제공됩니다.

IPv6는 현재 인터넷에서 점차 증가하고 있으며, 이전 버전인 IPv4와 함께 사용되는 것이 일반적입니다. IPv6는 인터넷의 미래를 위한 중요한 기술 중 하나로 간주됩니다.

20. 네트워크 구조

20.1. 네트워크 토폴로지

네트워크 토폴로지란, 네트워크에서 컴퓨터나 장치들이 연결된 형태를 말합니다. 이는 네트워크의 물리적인 구성을 나타내는 것으로, 네트워크의 속도, 안정성, 보안 등에 영향을 미칩니다.

네트워크 토폴로지의 종류는 다양하지만, 주요한 네트워크 토폴로지는 다음과 같습니다.

① 버스 토폴로지(Bus Topology)

모든 컴퓨터와 장치가 하나의 케이블에 연결되는 구조입니다. 버스 토폴로지는 구조가 단순하고 비용이 적게 들어가지만, 케이블의 손상으로 인한 네트워크 전체의 문제가 발생할 수 있습니다.

② 스타 토폴로지(Star Topology)

모든 컴퓨터와 장치가 중앙의 허브나 스위치에 연결되는 구조입니다. 스타 토폴로지는 버스 토폴로지와 비교하여 안정성이 높고 유지보수가 용이합니다.

③ 링 토폴로지(Ring Topology)

컴퓨터와 장치들이 순환 형태로 연결되어 있는 구조입니다. 링 토폴로지는 버스 토폴로지와 같이 케이블의 손상으로 인한 전체 네트워크 장애를 예방할 수 없지만, 성능이 우수하며 유지보수가 용이합니다.

④ 메시 토폴로지(Mesh Topology)

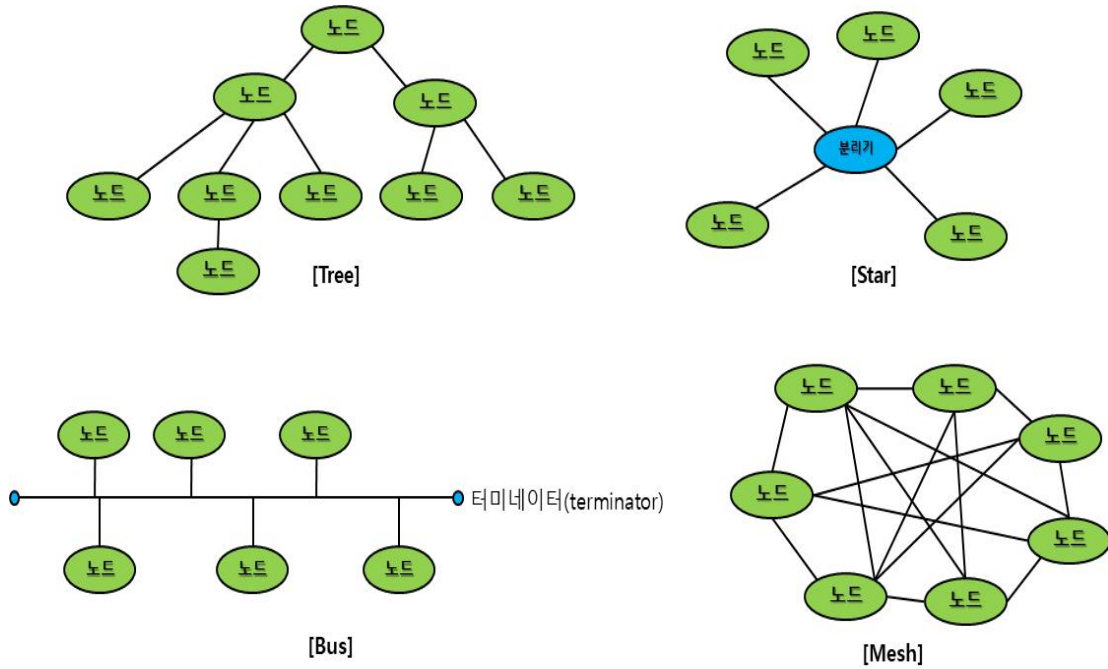
모든 컴퓨터와 장치들이 서로 직접적으로 연결되어 있는 구조입니다. 메시 토폴로지는 전송 속도가 빠르고 안정적이지만, 설치 비용이 매우 높아서 일반적으로 사용되지 않습니다.

⑤ 트리 토폴로지(Tree Topology)

스타 토폴로지의 확장형으로, 스타 토폴로지를 계층적으로 구성한 구조입니다. 높은 수준의 안정성과 유연성을 제공하지만, 전체 네트워크 성능이 중간 노드의 처리 능력에 따라 영향을 받을 수 있습니다.

⑥ 하이브리드 토폴로지(Hybrid Topology)

두 개 이상의 다른 토폴로지를 결합한 형태입니다. 예를 들어, 스타-버스 형태나 스타-링 형태 등이 있습니다.



[그림 3-36] 네트워크 토폴로지

20.2. 네트워크 분류

네트워크는 주로 지역적으로 LAN, MAN, WAN으로 구분하다.

LAN(Local Area Network)은 한정된 지리적 범위 내에서 컴퓨터와 기타 디바이스들이 연결되어 있는 컴퓨터 네트워크입니다. 일반적으로 집, 사무실, 학교, 공장 등의 건물 내부나 근거리에서 사용됩니다. LAN은 여러 대의 컴퓨터와 디바이스들을 물리적으로 연결하여, 서로 정보를 공유하고 데이터를 전송할 수 있도록 합니다.

LAN의 주요 특징은 다음과 같습니다.

① 지리적 한계

LAN은 지리적으로 제한된 작은 지역에 한정되어 있습니다. 대개 몇 백 미터에서 수백 킬로미터 이내의 범위에서 사용됩니다.

② 높은 전송 속도

LAN은 일반적으로 높은 전송 속도를 제공합니다. 이는 전용 선으로 연결되어 있거나 고속 무선 기술을 사용하기 때문입니다.

③ 저렴한 비용

LAN은 비교적 저렴한 비용으로 구성할 수 있습니다. 대부분의 경우, 간단한 이더넷 케이블과 네트워크 인터페이스 카드만 있으면 구성할 수 있습니다.

④ 공유 자원

LAN은 여러 대의 컴퓨터와 기타 디바이스들이 연결되어 있어, 프린터, 파일, 데이터베이스 등과 같은 자원을 공유할 수 있습니다.

⑤ 쉬운 관리

LAN은 중앙 집중식으로 관리될 수 있습니다. 네트워크 관리자는 네트워크를 모니터링하고, 사용자의 계정을 관리하고, 보안 정책을 적용하고, 소프트웨어 및 하드웨어 업데이트를 수행할 수 있으며, 주로 하나의 회사나 기관에서 관리를 하고 있습니다.

⑥ 보안

LAN은 일반적으로 높은 보안 기능을 제공합니다. 사용자 인증, 데이터 암호화 등의 보안 기능을 사용하여 불법적인 액세스나 데이터 유출을 방지할 수 있습니다.

⑦ 유연성

LAN은 다양한 디바이스와 프로토콜을 지원할 수 있습니다. 따라서, 다양한 유형의 기기를 추가하거나 기존의 시스템을 업그레이드할 때도 유연하게 대처할 수 있습니다.

이러한 특징들은 LAN을 일상적인 컴퓨터 네트워크로 만듭니다. 현재 대부분의 조직에서는 LAN을 중심으로 IT 시스템을 구축하고 있다.

두 번째로, MAN(Metropolitan Area Network)은 LAN보다는 넓은 지리적 범위를 가지는 컴퓨터 네트워크로, 도시나 대도시와 같이 상대적으로 넓은 지역에 위치한 여러 건물을 연결합니다. MAN은 여러 개의 LAN을 묶어서 하나의 대규모 네트워크를 형성합니다.

MAN의 주요 특징은 다음과 같습니다.

① 지리적 한계

MAN은 LAN보다 넓은 지역을 커버합니다. 보통 수십 킬로미터 이내의 지역에 사용됩니다.

② 중간 규모

LAN보다 크고 WAN보다 작은 규모의 네트워크입니다.

③ 높은 전송 속도

MAN은 일반적으로 높은 전송 속도를 제공합니다. 이는 광섬유 케이블이나 무선 기술을 사용하기 때문입니다.

④ 다양한 서비스

MAN은 다양한 서비스를 제공할 수 있습니다. VoIP(음성 통화), IPTV(인터넷 방송), 대용량 데이터 전송 등의 서비스를 제공할 수 있습니다.

⑤ 보안

MAN은 LAN보다는 보안이 취약한 편입니다. 이는 여러 건물을 연결하다 보니 보안에 취약한 지점이 늘어나기 때문입니다.

⑥ 비용

MAN은 LAN보다는 비용이 높을 수 있습니다. 여러 건물을 연결하기 위해서는 높은 비용이 필요하기 때문입니다.

⑦ 중앙 집중식 관리

MAN은 일반적으로 중앙 집중식으로 관리됩니다. 네트워크 관리자는 네트워크를 모니터링하고, 사용자의 계정을 관리하고, 보안 정책을 적용하고, 소프트웨어 및 하드웨어 업데이트를 수행할 수 있습니다.

이러한 특징들은 MAN을 대도시와 같이 넓은 지역을 커버해야 하는 조직에서 주로 사용됩니다. 예를 들어, 대학교, 국가 기관, 대규모 회사 등이 MAN을 구성하는 경우가 많습니다.

마지막으로, WAN(Wide Area Network)은 지리적으로 넓은 지역에 분산된 컴퓨터 네트워크를 연결하는 기술입니다. 인터넷이 가장 큰 WAN이며, 전 세계적으로 컴퓨터와 장치를 연결하는 대규모 네트워크입니다.

WAN의 주요 특징은 다음과 같습니다.

① 지리적 한계

WAN은 LAN과 MAN보다 넓은 지역을 커버합니다. 지리적으로 떨어져 있는 지역에 위치한 컴퓨터나 장치를 연결합니다.

② 대규모 네트워크

WAN은 대규모 네트워크입니다. 전 세계적으로 수많은 컴퓨터와 장치를 연결하고 있습니다.

③ 전송 속도

WAN은 LAN과 MAN에 비해 느릴 수 있습니다. 이는 지리적으로 떨어져 있는 장소를 연결해야 하기 때문입니다. 그러나 최근에는 광섬유 기술 등 빠른 전송 속도를 제공하는 기술이 발전하면서 속도가 향상되고 있습니다.

④ 다양한 기술

WAN은 다양한 기술을 사용합니다. 전화선, 광섬유, 무선, 위성 등 다양한 기술을 결합해서 사용합니다.

⑤ 공용 인프라

WAN은 대개 공용 인프라를 사용합니다. 인터넷은 전 세계적으로 공용 인프라를 사용하는 대표적인 WAN입니다.

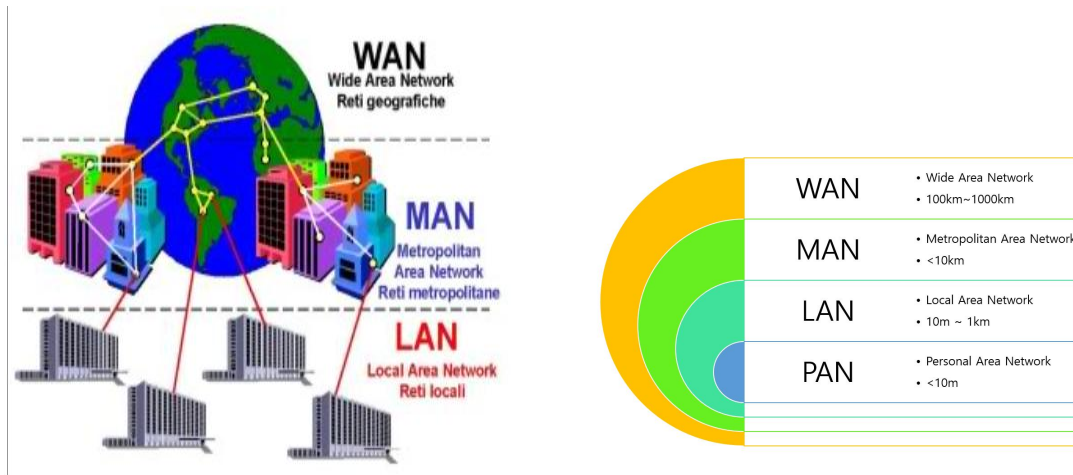
⑥ 보안

WAN은 LAN보다 보안이 취약할 수 있습니다. 인터넷과 같은 공용 인프라를 사용하다 보니 보안 위협이 존재할 수 있습니다.

⑦ 높은 비용

WAN 구축 비용은 높을 수 있습니다. 인터넷을 사용하는 경우 인터넷 서비스 제공자(ISP)에 대한 비용이 들어갑니다.

WAN은 전 세계적으로 컴퓨터와 장치를 연결하는 대규모 네트워크입니다. 이러한 특징들은 WAN을 대규모 기업, 금융 기관, 국가 기관 등에서 주로 사용됩니다.



[그림 3-37] 거리별 네트워크 분류(LAN, MAN, WAN)

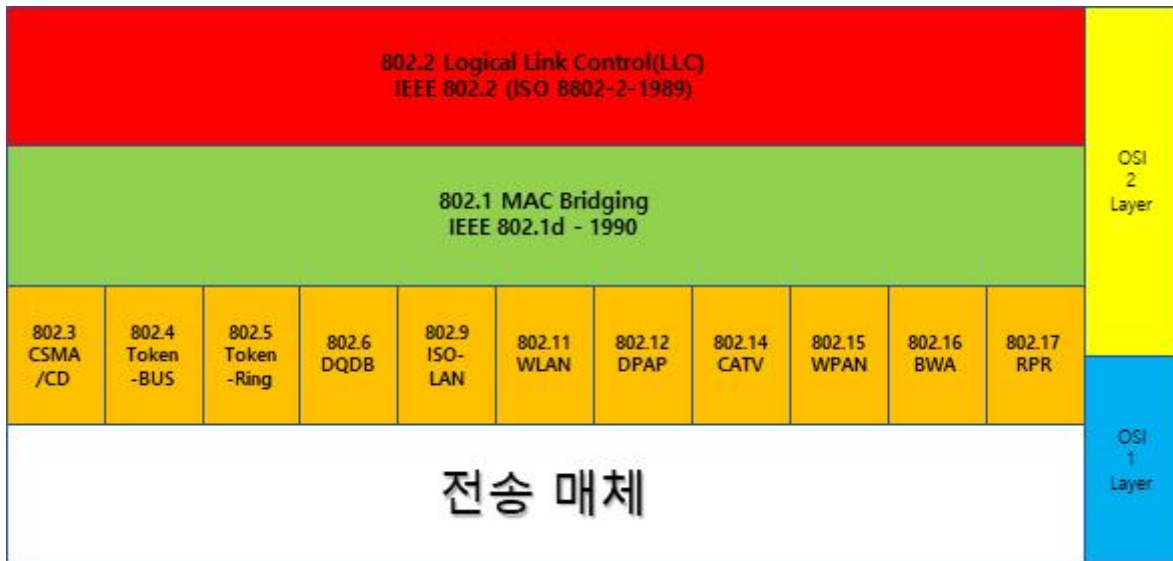
20.3. LAN 표준안

미국 전기 전자 학회(IEEE) 산하에서 근거리 통신망(LAN: Local Area Network)과 도시권 통신망(MAN: Metropolitan Area Network) 표준을 담당하는 IEEE 802 위원회에서 제정된 일련의 표준. IEEE 802.1, 802.2, 802.12 등이 있다.

IEEE 802 표준 LAN 프로토콜 참조 모델은 OSI 기본 참조 모델의 계층화 개념을 바탕으로 하고 있는데, 주로 OSI 모델의 7계층 가운데 하위의 2계층까지를 표준화하였다. 그러나 IEEE 802 표준 LAN의 데이터 링크 계층은 논리 회선 제어 부계층(LLC sublayer)과 매체 접근 제어 부분층(MAC sublayer)의 2개 서브 계층으로 나뉜다.

논리 회선 제어 부계층의 표준은 모든 IEEE 802 표준에 적용되는데, 데이터국(노드) 간 연결, 메시지 프레임 생성 및 오류 제어 기능을 수행한다. 매체 접근 제어 부분층의 표준은 매체 접근 제어 방식과 충돌 탐지(collision detection) 방식별로 규정되어 있다.

그리고 물리 계층에는 각 표준에 대한 다양한 전송 매체의 규격이 제시되어 있다. OSI 기본 참조 모델에서 전송 매체는 물리 계층의 아래에 존재하는 것으로 간주되지만, LAN의 설계에서 전송 매체의 선택이 중요하기 때문에 전송 매체의 규격이 물리 계층에 포함되어 있다. IEEE 802 LAN 표준을 도표로 표시하면 그림과 같다.



[그림 3-38] 802 표준안

802.1: 상위계층 인터페이스와 MAC BRIDGE

802.2: LLC(Logical Link Control)

802.3: CSMA/CD(Carrier Sense Multiple Access/Collision Detection)

802.4: 토큰버스(Token Bus)

802.5: 토큰링(Token Ring)

802.6: MAN(Metropolitan Area Network)

802.7: 광대역(Broadband) LAN

802.8: 광섬유(Fiber Optic) LAN

802.9: 종합데이터 & 음성 네트워크

802.10: 보안(Security)

802.11: 무선 네트워크(Wireless Network)

20.4. 무선 네트워크

무선 네트워크는 물리적인 케이블 대신 무선 신호를 사용하여 기기들을 연결하는 네트워크입니다. 이러한 무선 기술을 이용하여 기기 간의 데이터 전송이 이루어지며, 일반적으로 블루투스, Wi-Fi, NFC, 인프래레드 등의 기술이 사용됩니다.

① 블루투스(Bluetooth)

- 블루투스는 주로 개인적인 사용을 목적으로 하는 PAN(Personal Area Network)에 사용되는 무선 통신 기술입니다.
- 블루투스는 주로 스마트폰, 스피커, 이어폰, 스마트 워치 등과 같은 기기 간에 사용됩니다.
- 블루투스의 전송 속도는 빠르지는 않지만, 저전력 무선 통신 기술을 사용하기 때문에 배터리 수명이 오래간다는 장점이 있습니다.

② Wi-Fi

- Wi-Fi는 IEEE 802.11 표준을 기반으로 한 무선 네트워크 기술입니다.
- Wi-Fi를 사용하면 노트북, 스마트폰, 태블릿 등 다양한 기기에서 인터넷에 접속할 수 있습니다.
- Wi-Fi는 높은 전송 속도를 제공하며, 대부분의 가정, 사무실, 카페 등에서 사용됩니다.

③ NFC(Near Field Communication)

- NFC는 매우 가까운 거리에서 데이터를 전송하는 기술입니다.
- NFC는 스마트폰, 카드 판독기, 태그 등과 같은 기기 간에 사용됩니다.
- NFC는 결제, 신분증 인증, 교통카드 등 다양한 용도로 사용됩니다.

④ 인프래레드(Infrared)

- 인프래레드는 빛의 파장을 이용하여 데이터를 전송하는 기술입니다.
- 인프래레드는 매우 가까운 거리에서만 사용 가능하며, 전송 속도가 느리다는 단점이 있습니다.
- 인프래레드는 TV 리모컨, 스마트폰 등에서 사용됩니다.

⑤ Zigbee

- IEEE 802.15.4 표준을 기반으로 한 무선 네트워크 기술입니다.
- Zigbee는 저전력 무선 통신 기술로, 스마트 홈이나 스마트 빌딩 등의 IoT (Internet of Things) 기기들 간의 통신에 사용됩니다.

⑥ LTE (Long-Term Evolution)

- LTE는 모바일 무선 통신 기술로, 고속으로 데이터를 전송할 수 있는 기술입니다.
- LTE는 4G 네트워크로도 알려져 있으며, 모바일 디바이스에서 사용됩니다.

이러한 무선 기술들은 각각의 장단점이 있으며, 적절한 상황에서 사용해야 합니다. 무선 기술은 기존의 유선 기술과는 다른 특징을 가지고 있으며, 무선 환경에서의 보안 문제에 대한 고민도 필요합니다.