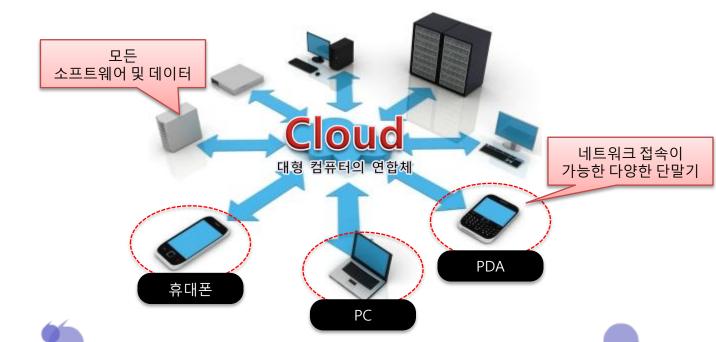
실무에 적용 가능한 Big Data 분석 개론

빅데이터 IT 이론



1. 클라우드 컴퓨팅의 개념

클라우드 컴퓨팅이란?



장소에 구애 받지 않고 원하는 작업을 수행할 수 있는 컴퓨팅 기술

1. 클라우드 컴퓨팅의 개념

클라우드 컴퓨팅의 다양한 정의

클라우드 컴퓨팅이란 대형 전력회사에서 전기를 받아쓰듯이 중앙 집중화된 대형 데이터센터에서 서비스를 받고 소프트웨어 프로그램들도 인터넷 망을 통해서 자유롭게 빌려 쓰며, IT 자원을 구매하거나 소유할 필요 없이, 필요한 만큼 사용료를 주고 쓰는 개념이다.





클라우드 컴퓨팅이란 개인컴퓨터가 아닌 인터넷과 연결된 메인 컴퓨터에 저장해 놓고 인터넷만 접속해 있으면 어떤 단말기로도 원하는 문서작업이 가능한 환경이다.

IBM Watson

클라우드 컴퓨팅이란 웹 기반 어플리케이션을 활용하여 대용량 데이터베이스를 인터넷 가상 공간에서 분산 처리하고, 이 데이터를 PC, 휴대폰, 노트북, PDA 등 다양한 단말기에서 불러 오거나 가공할 수 있게 하는 환경

1. 클라우드 컴퓨팅의 개념

클라우드 컴퓨팅 기술의 발전

자원 및 서비스의 아웃소싱이 확대

- ■분업화와 규모의 경제가 실현되고 있음
- 컴퓨팅 자원에 대한 소유방식의 변화가 나타남

온라인 서비스화

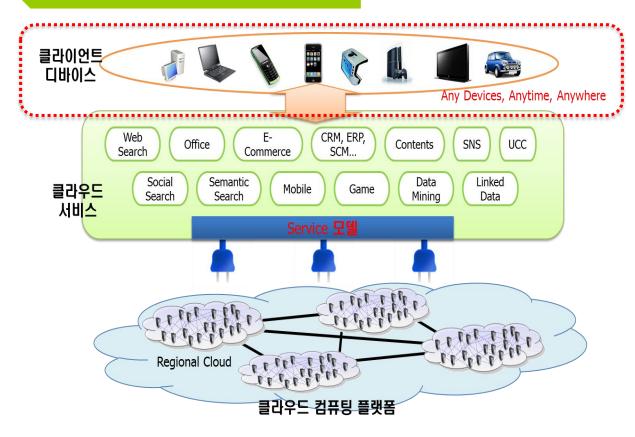
- 인터넷을 기반함
- ■소프트웨어와 콘텐츠의 온라인 서비스화가 추진되고 있음

안정적인 IT 서비스 전송

■ 초고속망을 통하여 안정적인 IT 서비스 전송이 가능하게 됨

1. 클라우드 컴퓨팅의 개념

클라우드 컴퓨팅 개념도

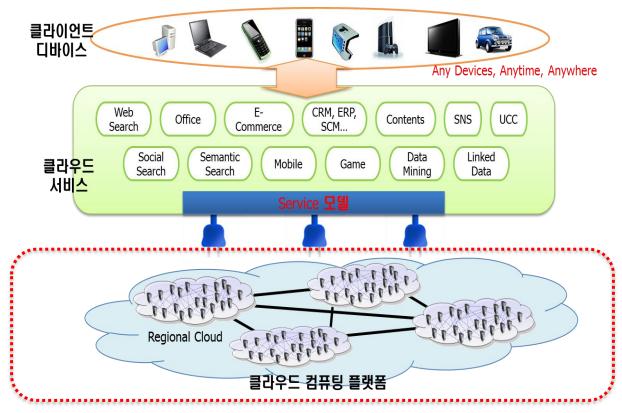


▶ 클라이언트 디바이스

- 최종 사용자가 인터넷에 올라오는 정보를 관리할 수 있도록 사용하는 모든 기기
- 일반 컴퓨터, 랩톱 컴퓨터, 태블릿 PC, 휴대전화, PDA 등의 기기들이 클라이언트 디바이스에 해당함
- 클라이언트 디바이스를 줄여서 클라이언트라고 함

1. 클라우드 컴퓨팅의 개념

클라우드 컴퓨팅 개념도

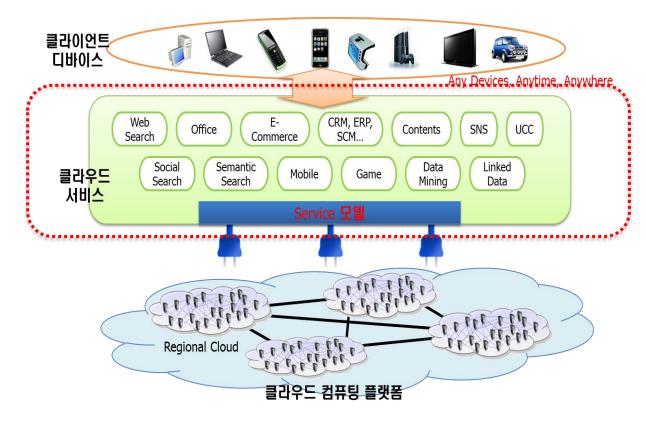


▶ 클라이언트 컴퓨팅 플랫폼

- 클라이언트에서 사용하는 어플리케이션이 설치된 서버인 데이터센터가 있음
- 서버를 분산한 분산 서버로 구축됨 → 한 곳에 있는 것처럼 느껴짐
- 서버가 더 필요한 경우 다른 곳의 물리서버에 가상 서버를 추가하면 됨

1. 클라우드 컴퓨팅의 개념

클라우드 컴퓨팅 개념도



▶ 클라우드 서비스

■ 서비스 형태로 어플리케이션(CRM, ERP 등)을 제공함

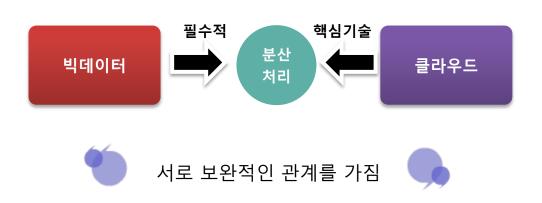
고객관계관리시스템

전사적 자원 관리

- 어플리케이션과 서비스를 제공받도록 모든 자원을 공급함
- 고객의 요구사항을 처리할 수 있도록 하드웨어를 제공함

2. 빅데이터와 클라우드 컴퓨팅

빅데이터와 클라우드의 관계



클라우드 서비스를 제공하는 대표 기업

인터넷 검색 서비스가 핵심

구글 앱스(Google Apps)

- 개인 클라우드 서비스
- ■이메일 기능(Gmail)
- 문서도구(Google Docs)
- ■데이터 연산기능 (Goggle Spreadsheet)

빅 쿼리(Big Query) 소프트웨어

- 빅데이터 분석소프트웨어
- ■테라바이트급의 대용량 데이터를 구글 검색엔진 인프라로 실시간 분석함

구글 예측 프로그램 (Google Prediction API)

- 예측 프로그램
- ■기계학습을 통해 일정한 데이터 패턴을 발견하고 이를 통해 새로운 예측결과를 제공함
 - 예 스팸메일 자동 삭제, 자동차 운행 경로 찾기 등

2. 빅데이터와 클라우드 컴퓨팅

클라우드 서비스를 제공하는 대표 기업

amazon.com.

인터넷 서점으로 출발한 기업

서적, 상품저장을 저장하기 위해 대량의 서버와 데이터베이스 구축

최대치를 기준으로 설계되어 다른 기업에게 서비스로 제공할 수 있게 됨

S3 (Simple Storage Service)

> 저장 장치를 빌려줌

심플 DB (Simple DB)

데이터베이스를 빌려줌 EC2 (Elastic Computing Cloud)

서버를 빌려줌

EMR (Elastic Map Reduce)

빅데이터 분석용 프로그램



컴퓨터에 운영체제(OS) 설치하는 사업 클레우드와 정반대 사업

윈도우 애저 (Window Azure) 서비스

- 기존의 사업모델을 파괴하는 것이 아님
- 윈도우 기반과 클라우드 기반을 연계하는 혼합형 서비스를 제공함

실무에 적용 가능한 Big Data 분석 개론

빅데이터 IT 이론



1. 데이터베이스와 데이터베이스 관리 시스템

데이터베이스(DB)란?

정의

어느 한 조직의 다양한 응용 프로그램들이 공동으로 사용하는 데이터들을 통합하여 저장한 운영 데이터의 집합

특징

통합된 데이터

- 여러 부서에서 사용하는 데이터를 한 곳에 모아서 공동으로 관리함
- ■최소한의 중복만 허용함

저장된 데이터

■컴퓨터가 접근할 수 있는 디스크와 같은 저장 매체에 저장된 데이터

운영 데이터

■조직의 운영에 꼭 필요한 데이터를 저장한 운영 데이터

데이터베이스 관리 시스템 (DBMS, DataBase Management System)



데이터베이스 관리 시스템이란?

사용자와 응용 프로그램에 편리하고 효율적인 데이터베이스 사용 환경을 제공하는 소프트웨어

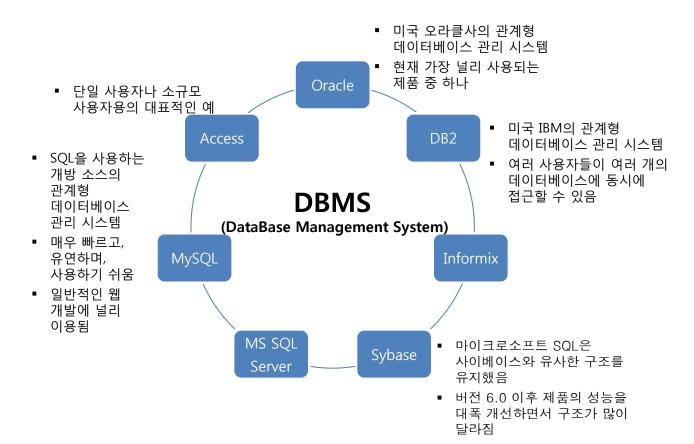


사용목적

- 데이터베이스의 구조를 변경해도 응용 프로그램에 영향을 주지 않고, 반대로 응용 프로그램을 변경해도 데이터베이스에 영향을 주지 않도록 하는 것
- ▶ 응용 프로그램이 데이터에 종속되지 않는 데이터 독립성을 제공하는 것

1. 데이터베이스와 데이터베이스 관리 시스템

데이터베이스 관리 시스템 (DBMS, DataBase Management System)



1. 데이터베이스와 데이터베이스 관리 시스템

데이터베이스 관리 시스템 (DBMS, DataBase Management System)

관계형 데이터베이스 관리 시스템 (relational DBMS)

- 데이터베이스 모델 중 하나
- 데이터들을 2차원의 테이블들로 표현함
- 각각의 테이블은 하나의 데이터베이스화되는 대상인 개체들의 속성들과 관련한 데이터들로 구성됨

속성

학생 테이블

학번	성명	학과명	학년	연락처	이메일
06120001	홍길동	컴퓨터공학과	2	010-123-5678	hong@naver.com
07120010	김재철	전자공학과	3	011-2345-4321	kim@hanmail.net
08130021	박은영	정보통신공학과	3	016-987-6543	park@nate.com
09140101	홍길동	의용공학과	2	019-456-7890	gildong@yahoo.com

<출처: IT CookBook, 컴퓨터 사이언스: 기본 컴퓨터 이론부터 최신 정보통신 기술까지>

1. 데이터베이스와 데이터베이스 관리 시스템

데이터베이스 관리 시스템의 기능

데이터 정의 기능

데이터 조작어 기능

리포트 생성 기능

- 데이터베이스의 테이블들을 생성하고 각 테이블의 속성들의 특성 들을 정의하여 데이<mark>터베이스의 콘텐츠의 구조를 규정하는</mark> 기능을 의미함
- 데이터 사전
 - ✓ 데이터베이스에 대한 정보임
 - ✓ 데이터 요소의 정의와 특성을 저장한 자동화되고 수동적인 파일

데이터 정의 기능

데이터 조작어 기능

리포트 생성 기능

- 데이터베이스에서 데이터를 추가하고, 변경하고, 삭제하고, 조회하는 등의 기능
 - ✓ 데이터의 검색, 삽입, 삭제, 갱신 등과 같은 데이터베이스 연산을 처리하는 기능 제공
- SQL(Structured Query Language)은 관계형 데이터베이스의 조작과 관리에 사용되는 데이터베이스 질의용 언어
 - ✓ 다수의 DBMS에서 구조화된 질의어로 널리 사용되고 있음

가격이 30,000원 이하인 도서를 검색하는 SQL 예

SELECT title_id, title, price **FROM** BookTable

WHERE price <=30000;

<출처: IT CookBook, 컴퓨터 사이언스: 기본 컴퓨터 이론부터 최신 정보통신 기술까지>

1. 데이터베이스와 데이터베이스 관리 시스템

데이터베이스 관리 시스템의 기능

데이터 정의 기능

데이터 조작어 기능

리포트 생성 기능

- 데이터베이스를 활용한 세련된 리포트를 생성함
- 크리스탈 리포트
 - ✓ 마이크로소프트 액세스에서도 사용할 수 있으나 대기업용 데이터베이스 관리 시스템 리포트 생성기로 잘 알려져 있음

2. NoSQL

새로운 저장 기술 등장

인터넷과 컴퓨터의 발전으로 많은 양의 데이터가 수집 가능해짐

비정형 데이터의 처리 등에 어려움이 생김

데이터 저장 구조가 한계에 달함



NoSQL이 등장함



기존 관계형 데이터베이스 관리 시스템과 NoSQL

기존 관계형 데이터베이스 관리 시스템

- · 일관성과 가용성이 보장됨
- ✓ 일관성 : 분산 환경에서 모든 노드가 같은 시점에 같은 데이터를 보여줘야 함
- ✓ 가용성 : 일부 노드가 다운되어도 다른 노드에 영향을 주지 않아야 함
- 클라우드 분산 환경에 부적합함
 - ✓ 대규모 데이터 처리에 운영 환경, 기능, 성능면에서 확장성 혹은 고비용 문제

NoSQL

- 지속성을 보장함
- ✓ 지속성 : 성공적으로 수행된 트랜잭션은 영원히 반영되어야 함
- ✓ 일관성이나 가용성 중 하나를 포기함
- ✓ 네트워크 전송 중 일부 데이터를 손실하더라도 시스템은 정상 동작하는 것을 보장함
- 웹 환경의 다양한 정보를 검색하고, 저장하는 것이 가능함

노드 : 네트워크의 기본요소인 지역 네트워크에 연결된 컴퓨터와 그 안에 속한 장비들을 통틀어 지칭함

VS.