



# 통합 구현

## 대용량 서비스 레퍼런스 아키텍처 3



한국기술교육대학교  
온라인평생교육원

## ◆ 학습내용 ◆

- > Analysis Layer
- > OAM Layer
- > 클라우드 서비스

## ◆ 학습목표 ◆

- > Analysis Layer의 역할을 이해하고 분석 시스템의 구현방안을 설명할 수 있다.
- > OAM Layer의 역할을 이해하고, 관리 시스템을 활용할 수 있다.
- > 클라우드 컴퓨팅을 정의하고, 클라우드 서비스를 분별하여 적용할 수 있다.



# Analysis Layer

## 1. Analysis Layer

### 1) 개요

#### (1) 정의 및 용어

##### ① 정의

#### Analysis Layer란

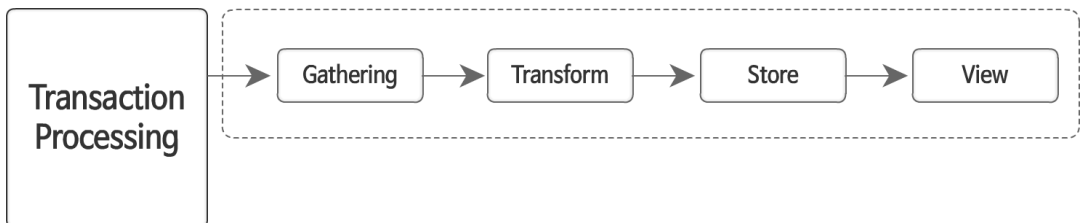
- 트랜잭션 처리에 의한 결과와 로그를 분석하는 계층

##### ② 용어

- Analysis Layer, BSS(Business Support System), 정보계(금융권) 등의 용어로 사용

#### (2) 구조

- Gathering 컴포넌트 : 트랜잭션 프로세싱 계층에서 생성된 로그와 데이터를 수집
- Transform 컴포넌트 : 분석을 위한 저장소에 저장하고자 수집된 로그와 데이터의 포맷 변경
- Store 컴포넌트 : 변경된 자료를 정제된 형태로 저장
- View 컴포넌트 : 리포트나 대시 보드 형태로 출력



[데이터 분석 계층의 구조]



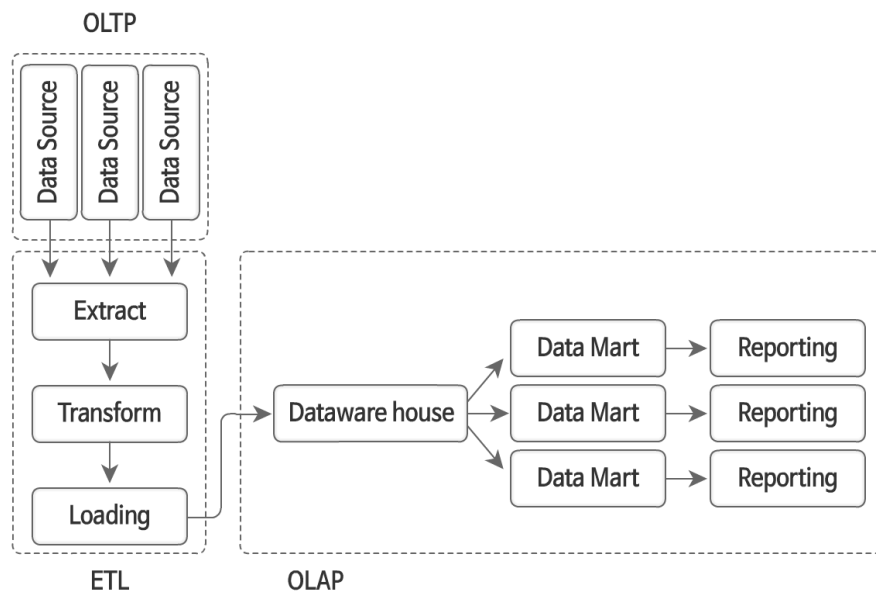
## Analysis Layer

### 1. Analysis Layer

#### 2) OLAP 방식

##### OLAP 방식이란?

- OnLine Analytic Processing
- RDBMS 기반의 분석 시스템
- 분석에 최적화된 데이터베이스를 사용하여 데이터분석 및 리포트 생성



[OLAP 기반의 데이터 분석 구조]



## Analysis Layer

### 1. Analysis Layer

#### 2) OLAP 방식

##### (1) ETL(Extract Transform Loading)

- ① 여러 데이터 소스로부터 수집해 변환 및 저장하는 역할
  - Extract : 다양한 데이터 소스로부터 데이터를 추출
  - Transform : 추출된 데이터를 변환하는 역할, 필요한 데이터의 선별 추출, 필드 합치기나 특정 규칙에 따라 데이터 변환
  - Loading : 변환된 데이터를 수신 쪽 데이터베이스에 저장
- ② ETL 외 분산 컴퓨팅 환경에 전문화된 로그 수집 솔루션으로 Flume 등이 있음

##### (2) 데이터 웨어하우스와 데이터 마트

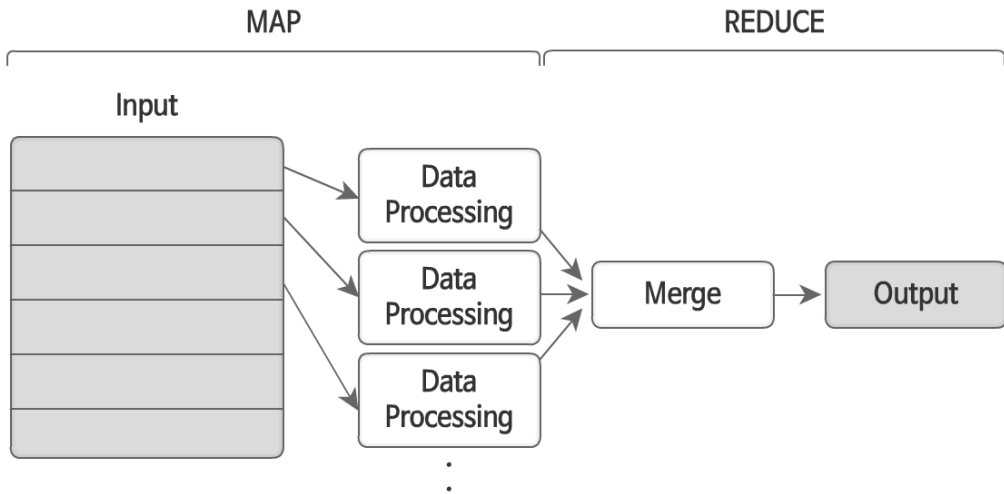
- ① 데이터 웨어하우스(Data Warehouse)
  - OLTP 시스템에서 수집된 데이터가 저장되는 중앙 데이터베이스
- ② 데이터 마트(Data Mart)
  - 리포트를 보고자 하는 각 부서에 따라 다시 정제되어 부서별로 저장되는 장소
  - 데이터 마트를 사용하는 경우 부서간 업무 성격에 따라 구조 변경이 용이하고, 각 데이터 저장 기간과 정책을 자유롭게 정할 수 있음
  - 기업 크기가 크지 않고, 데이터 분석을 필요로 하는 부서 수가 많지 않다면 별도 데이터 마트 없이 데이터 웨어하우스 하나로 대신할 수 있음



## Analysis Layer

### 1. Analysis Layer

#### 3) Map&Reduce 기반 분석시스템



#### [Map & Reduce의 기본 개념]

- 저가의 하드웨어로 대용량 분석 처리가 가능하나, 데이터 수집, 변환, 분석 등의 과정을 직접 구현해야 하는 프로그래밍 프레임워크
- 기존 방법으로는 대용량 데이터 분석이 어려움
- Map & Reduce 방법 활용
  - Map : 하나의 큰 데이터를 여러 개의 조각으로 나누어 처리하는 단계
  - Reduce : 처리 결과를 모아 하나로 합쳐 결과를 내는 단계
  - Hadoop(Apache)

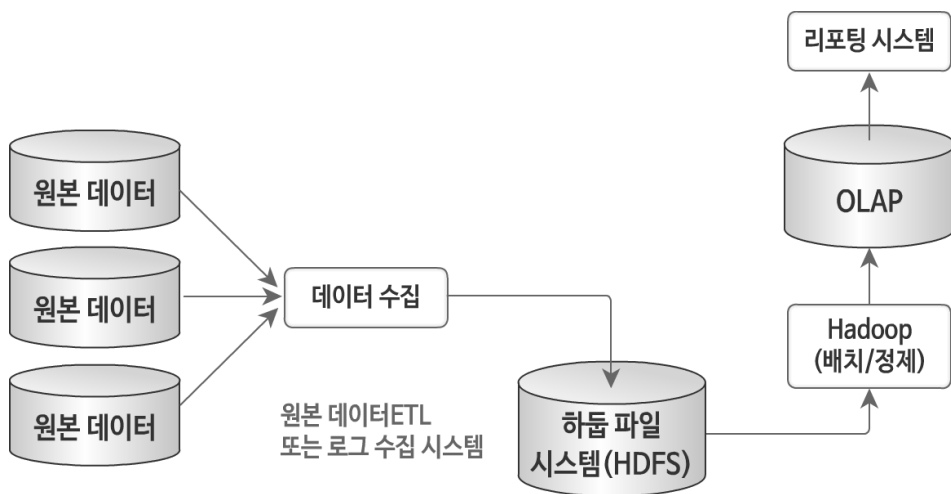


## Analysis Layer

### 1. Analysis Layer

#### 4) OLAP + Map&Reduce 기반

- 대용량 데이터 분석에 적합한 Map&Reduce 아키텍처와 유연한 데이터 리포팅이 가능한 OLAP 아키텍처를 조합해 사용



[Map & Reduce와 OLAP을 결합한 데이터 분석 구조]



## Analysis Layer

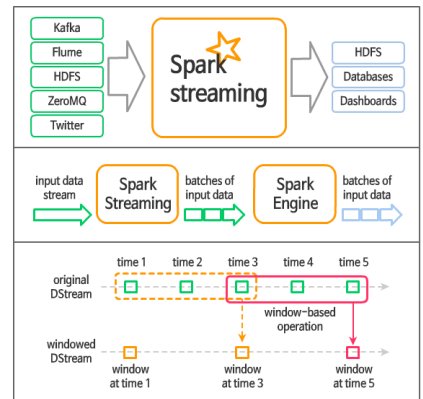
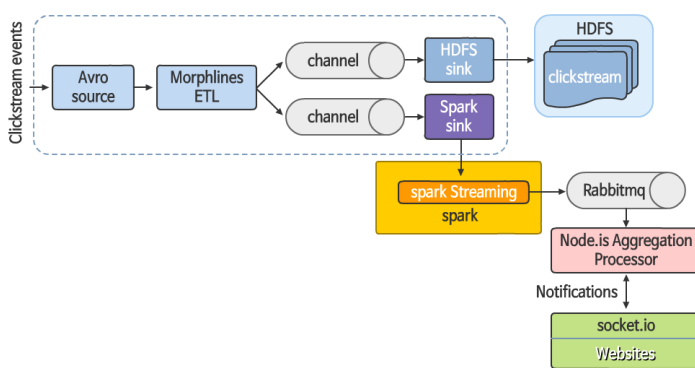
### 1. Analysis Layer

#### 5) 실시간 분석 시스템

- 전통적인 OLAP, Map&Reduce 방식은 배치 방식으로 데이터 생성 후 1~2일 후에 리포트 생성되어 신속한 의사결정이 어려움

#### 실시간 분석 시스템이란?

- 실시간으로 산출되는 데이터를 바로 처리할 수 있는 스트리밍 기술이 필요하며, 최근 오픈 소스인 Storm, Spark 등을 실시간 분석에 활용







## 2. OAM Layer

### 1) 개요

- OAM은 Operation, Administration & Monitor의 약자로 OSS(Operation Support System)로도 불림
- 서비스할 때 시스템의 운영자 입장에서 필요한 관리 및 모니터링 기능 제공
- 구분
  - CMDB(Configuration Management DB) : 설정 관리
  - Deployment 시스템 : 시스템 배포 관리
  - Monitoring 시스템

### 2) CMDB (Configuration Management DB)

#### CMDB란?

- 여러 인스턴스로 구성된 분산 시스템이나 여러 컴포넌트로 조합된 시스템을 구축할 때 각 컴포넌트 및 인스턴스에 대한 설정 정보를 저장하기 위한 공용 데이터베이스
- 톰캣, WebLogic은 XML 파일로 CMDB 정보 관리
- 오라클, MySQL은 DB 자체에 CMDB 정보 관리
- ZooKeeper
  - 분산 시스템에서 집중화된 설정 정보를 저장하고 관리하는 시스템으로 Auto Scale out(자동으로 인스턴스의 수를 조정) 기능, 설정 변경 값에 대한 이벤트 트리거 기능 등을 제공하여 분산시스템 조정(coordination)



## 2. OAM Layer

### 3) 모니터링

- 시스템의 건전성 확인 및 장애에 대한 전조 인식, 장애 발생시 이를 추적하는데 근거가 되는 데이터 제공

#### (1) 모니터링 대상 계층

##### ① Infrastructure

- 서버, 스위치, 디스크 등의 실제 하드웨어 인프라에 대한 모니터링 지원
- SNMP라는 표준 프로토콜이나 이를 확장한 모니터링 인터페이스 제공
- 오픈소스 도구 - Nagios, Cacti, Ganglia 등
- 상용 도구 - HP OpenView, Unicenter TNG

##### ② DBMS

- 데이터베이스에 대한 모니터링
- 대부분 데이터베이스 솔루션 벤더에서 제품 제공

##### ③ Middleware

- 미들웨어에 대한 모니터링
- 톰캣, WebLogic 등의 애플리케이션 서버나 Rabbit MQ 등의 메시지 큐 미들웨어
- 대부분 해당 솔루션에 모니터링 포함

##### ④ Application

- 애플리케이션 자체 모니터링
- 일반적으로 개발자가 해당 애플리케이션에 맞게 개발
- 애플리케이션 서버와 애플리케이션을 함께 모니터링해주는 APM(Application Performance Monitoring) 제품 사용
- 제니퍼소프트의 Jennifer는 국산 모니터링 제품으로 애플리케이션 서버, 실시간 동시 접속자 수, 애플리케이션 병목 구간 등을 분석



## 2. OAM Layer

### 4) 로그 관리

#### 로그란?

- 시스템에서 발생하는 모든 행위에 대한 기록 (빅데이터와 연관)

#### ① 로그의 분류

- 시스템 로그 : OS에서부터 미들웨어에서 올라오는 로그로 문제 발생시 장애 원인 파악 및 튜닝 위한 기초 자료 등으로 활용
- 애플리케이션 로그 : 개발자가 작성한 애플리케이션에서 나오는 로그로 애플리케이션의 장애 원인 파악이나 디버깅을 위해 생성
- 비즈니스 로그 : 사용자의 서비스 사용 형태, 거래 기록 등 애플리케이션을 활용하는 형태에 대해 인위적으로 만들어내고 수집하는 로그

#### ② 수집 및 저장

- Flume, Fluentd : 분산 로그 시스템
- 대용량 분산 큐(Kafka) : 로그를 한 곳에 수집해 HDFS 등의 대용량 파일 시스템에 저장
- Elastic Search : 검색을 활용해 최근 몇 개월간의 시스템 로그들을 수집, 저장
- FTP, Log4J 등 이용하여 로그 파일 직접 수집
- 로그 수집 프레임워크 LogStash

#### ③ 분석 및 시각화

- 데이터 분석 아키텍처 이용(OLAP, Storm/Spark 등)
- Saiku, BI 대시보드, kibana 등의 시각화 도구 활용
- <http://www.meteorite.bi/products/saiku>
- <https://powerbi.microsoft.com/ko-kr/documentation/powerbi-service-dashboards/>
- <https://www.elastic.co/kr/products/kibana>



## 2. OAM Layer

### 4) 로그 관리

#### (1) 시스템 로그에 대한 레퍼런스 아키텍처

- 시스템 로그와 애플리케이션 로그는 장애 해결을 위한 용도로 사용
- 비즈니스로그는 통계 분석 용도로 사용
- 애플리케이션로그는 이벤트(경고, error)성 로그로 이벤트를 감지해서 확인하고 해결해야 하는 형태의 로그
  - Sentry (<https://sentry.io/welcome/>)와 같은 실시간 오류 추적 솔루션이 도움을 줄 수 있음
- 시스템 로그는 모아서 저장하고 나중에 검색하거나 전체 로그의 추이(사용자 유입량, CPU 사용량 등)를 모니터링 할 때 사용하고, 일반적으로 Log Stash + Kibana + Elastic Search 조합으로 로그처리 시스템 구성



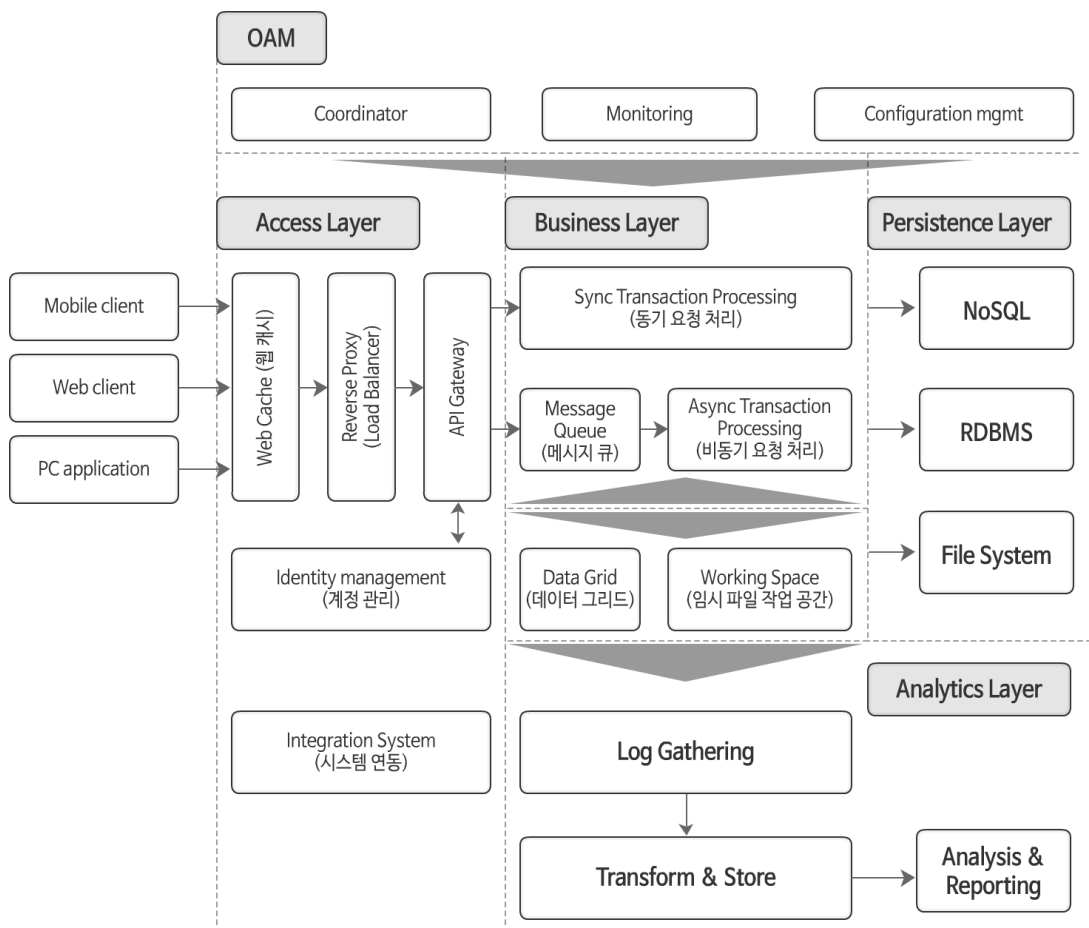
## OAM Layer

### 2. OAM Layer

#### 5) Configuration 관리

- 개발된 애플리케이션을 서버에 배포하는 모듈
- 클라우드 환경이 되며 Auto Scale out을 지원하면서 배포에 대한 요구 사항이 높아짐
- 자동화 도구 : Puppet, Chef, Ansible 등

#### 6) 대용량 서비스 레퍼런스 아키텍처





## 클라우드 인프라

### 3. 클라우드 인프라

#### 1) 클라우드 컴퓨팅

##### 클라우드 컴퓨팅이란?

- 서버나 네트워크, 스토리지와 같은 컴퓨팅 자원을 언제 어디서든 원격의 공유된 풀에서 필요한 경우 요청하여 필요한 만큼 사용하는 모델
- 기존 서비스에 비해 원격의 컴퓨팅 자원을 매우 빠르게 사용 가능
- 개인 클라우드(Public Cloud) - iCloud, DropBox, 네이버 N 드라이브 등
- 기업 클라우드(Enterprise Cloud) : 여기서는 기업 클라우드에 대한 내용 다룸

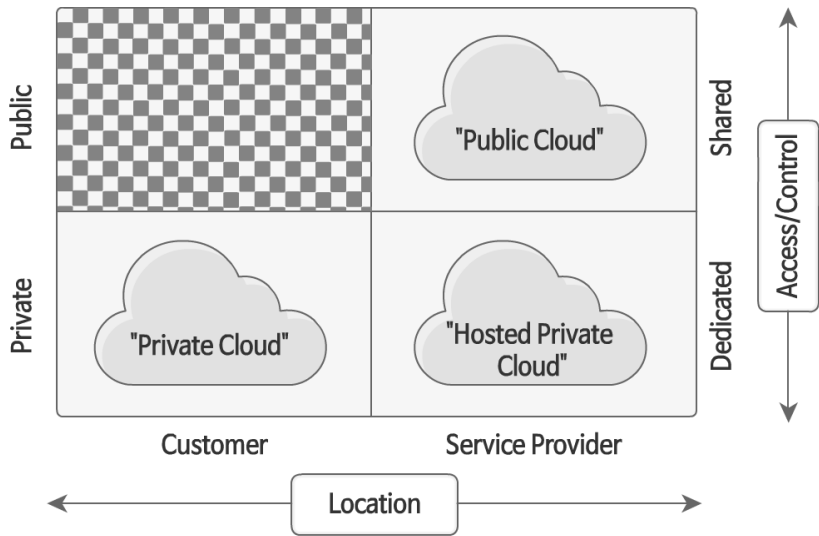


3. 클라우드 인프라

2) 배포 모델에 따른 분류

(1) 배포 장소 및 서비스 사용 제공/사용 주체에 따라 분리

- Private Cloud
- Public Cloud
- Hosted Private Cloud



[배포 모델에 따른 클라우드 컴퓨팅 분류]



### 3. 클라우드 인프라

#### 2) 배포 모델에 따른 분류

##### (1) 배포 장소 및 서비스 사용 제공/사용 주체에 따라 분리

###### ① Private Cloud

- 서비스 사용자가 기업 내부의 비즈니스 시스템을 위해 자체적으로 회사 내부 혹은 서드파티 데이터 센터에 클라우드 플랫폼을 구축하는 모델
- 예시
  - 데이터 센터에 Hosting Service + 가상화 이용
  - Hosting.com의 Cloud Dedicated Offering
  - Vmware, CloudStack 등

###### ② Public Cloud

- 서비스 제공자가 클라우드 서비스를 제공하기 위한 플랫폼
- 전문 클라우드 사업자(MS, Amazon 등)에 의해 서비스 제공
- 리소스 사용량에 따라 과금
- 예시
  - Amazon Web Service(AWS)
  - Windows Azure, Microsoft Dynamic Online, Office 365
  - Google Web Engine
  - IBM Softlayer, Bluemix

###### ③ Hosted Private Cloud

- 서비스 사용자가 기업 내부의 비즈니스 시스템을 서비스 제공자의 퍼블릭 클라우드 플랫폼을 인프라로 사용하여 구축하는 모델





# 클라우드 인프라

## 3. 클라우드 인프라

### 3) 서비스 단계에 따른 분류

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

- IT 서비스 제공 위한 주요 인프라 자원(CPU, 메모리, 디스크, 네트워크 환경 등)을 공유 자원 형태로 관리해 이를 나누어 제공하는 형태의 서비스
- 인프라의 리소스를 할당 받아 OS와 미들웨어를 설치해 서비스 이용
- 아마존의 AWS, MS의 Azure, IBM의 SoftLayer, 구글, RackSpace, HP 등 많은 클라우드 사업자가 있음

The screenshot shows the AWS Free Tier page. The header includes the AWS logo and navigation links. The main content area is titled 'AWS 프리 티어' (AWS Free Tier). It explains that the free tier allows users to try AWS services at no cost for 12 months. A list of services is provided, including Amazon EC2 and Amazon S3, with their respective free usage limits. A '지금 시작하기' (Get Started) button is located at the bottom of the page.

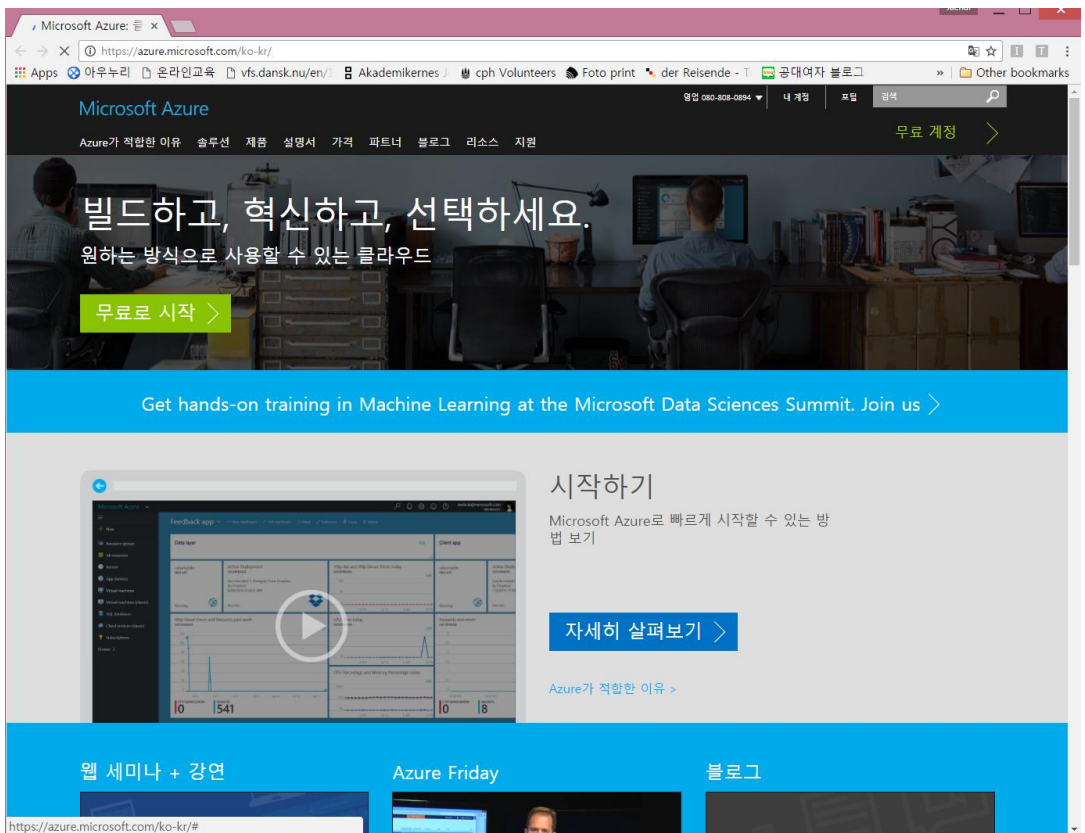


## 3. 클라우드 인프라

### 3) 서비스 단계에 따른 분류

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

- IaaS에 한 계층을 올려 소프트웨어를 개발할 수 있는 플랫폼 환경을 제공
- 개발 언어(Java, .NET, Rails, PHP, Node.js 등)가 구동될 수 있는 미들웨어 및 DB와 오픈 API 형태로 구현된 서비스 라이브러리 등을 제공
- 개발 트렌드가 모바일 앱이 되고, 스타트업이 중심이 되면서 적은 인원으로 빠르게 개발하고 관리가 가능한 플랫폼이 필요해져 PaaS가 주목받게 됨
- 구글의 AppEngine, MS Azure, Heroku, IBM의 Bluemix 등





### 3. 클라우드 인프라

#### 3) 서비스 단계에 따른 분류

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

- 이메일, CRM 등의 완성된 형태의 소프트웨어 서비스를 제공하는 형태의 클라우드 서비스
- ASP(Application Service Provider, Cafe24의 쇼핑몰 호스팅 등)과 유사한 서비스 모델
- 구글의 Gmail, Apps 서비스, Salesforce.com의 CRM 서비스, MS Office365 등



## 클라우드 인프라

### 3. 클라우드 인프라

#### 4) 클라우드 컴퓨팅의 장단점

##### (1) 장점

- 웹 콘솔을 사용해 설정 및 배포가 가능해 쉽고 빠름
- 초기 투자 비용이 저렴함
- 무제한적인 확장성 및 탄력성

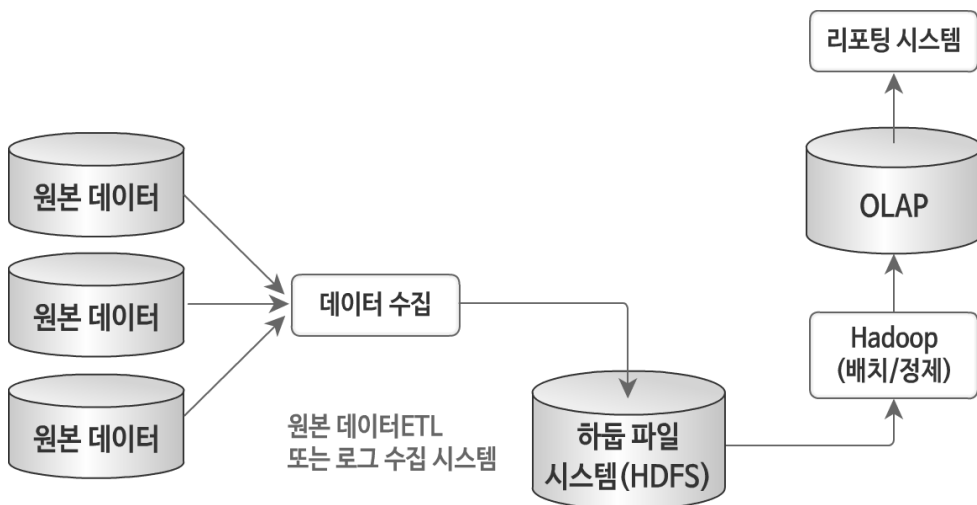
##### (2) 단점

- 서버 비용 외 네트워크, 디스크 등에 대한 비용이 추가
- 인프라를 다른 사용자와 공유해서 사용하면서 성능이 떨어짐
- 멀티 캐스트들이 대부분 지원되지 않아 다른 아키텍처를 가져야 함
- 다른 사용자와 물리적인 하드웨어를 공유하기 때문에 절대적인 성능을 예측할 수 없음
- VM이 사라지거나, 서비스 가용 시간이 100%가 아니기 때문에 불안정함

## ◆ 핵심정리 ◆

### 1. Analysis Layer

- 개요
  - 트랜잭션 처리에 의한 결과와 로그를 분석하는 계층
- OLAP 방식(OnLine Analytic Processing)
  - RDBMS 기반의 분석 시스템
- OLAP + MAP & Reduce 기반



## ◆ 핵심정리 ◆

### 2. OAM Layer

- 개요
  - Operation, Administration & Monitor
  - 시스템의 운영자 입장에서 필요한 관리 및 모니터링 기능 제공
- CMDB
  - 여러 인스턴스로 구성된 분산 시스템이나 여러 컴포넌트로 조합된 시스템을 구축할 때 각 컴포넌트 및 인스턴스에 대한 설정 정보를 저장하기 위한 공용 데이터베이스
- 모니터링
  - 시스템의 건전성 확인 및 장애에 대한 전조 인식, 장애 발생시 이를 추적하는데 근거가 되는 데이터 제공

## ◆ 핵심정리 ◆

### 3. 클라우드 인프라

- 클라우드 컴퓨팅
  - 서버나 네트워크, 스토리지와 같은 컴퓨팅 자원을 언제 어디서든 원격의 공유된 풀에서 필요한 경우 요청하여 필요한 만큼 사용하는 모델
- 배포 모델에 따른 분류
  - Public Cloud, Private Cloud, Hosted Private Cloud
- 서비스 단계에 따른 분류
  - IaaS(Infrastructure as a Service)
  - PaaS(Platform as a Service)
  - SaaS(Software as a Service)

