

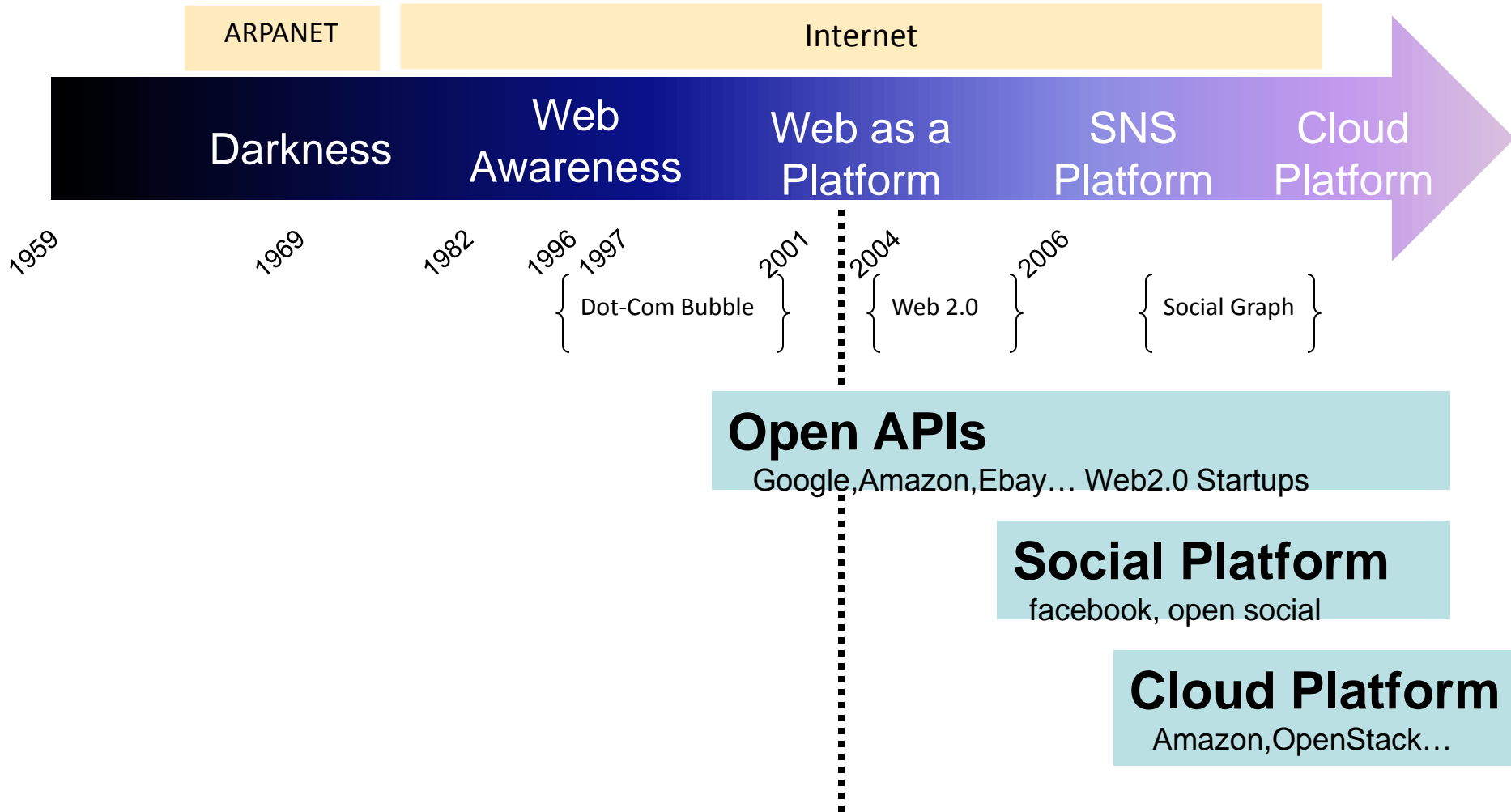


클라우드 컴퓨팅 서비스 현황 및 Daum의 사례

Channy Yun

Daum Communications Corp.
channy@daumcorp.com

웹 서비스 컴퓨팅 변천



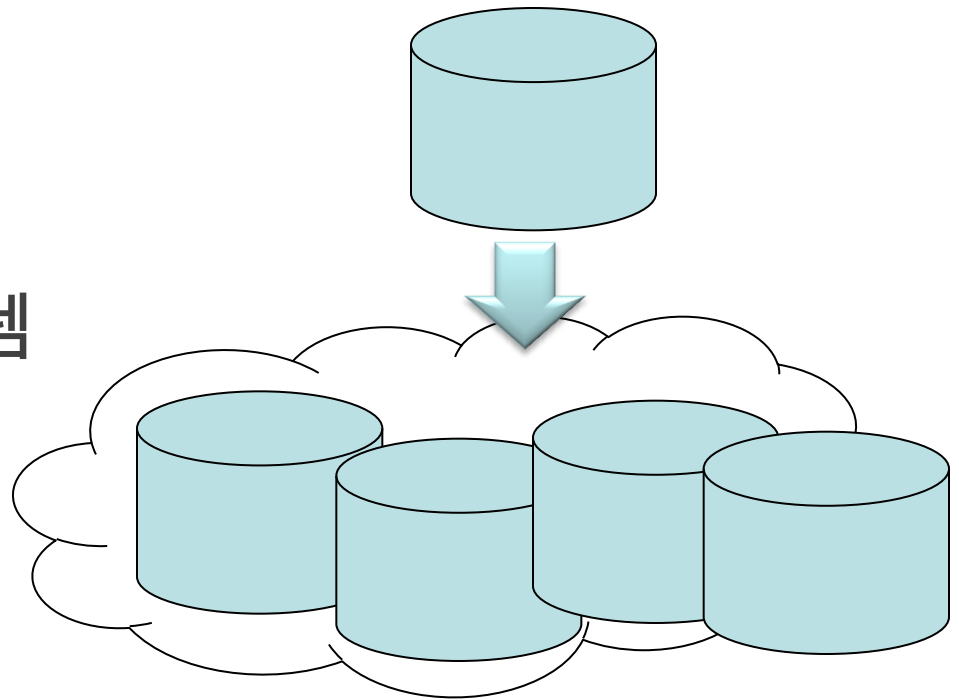
서비스 플랫폼의 변화

- 과거

- 단일한 고립 시스템
- 견고성에 집중
- Highly coupled 시스템

- 현재

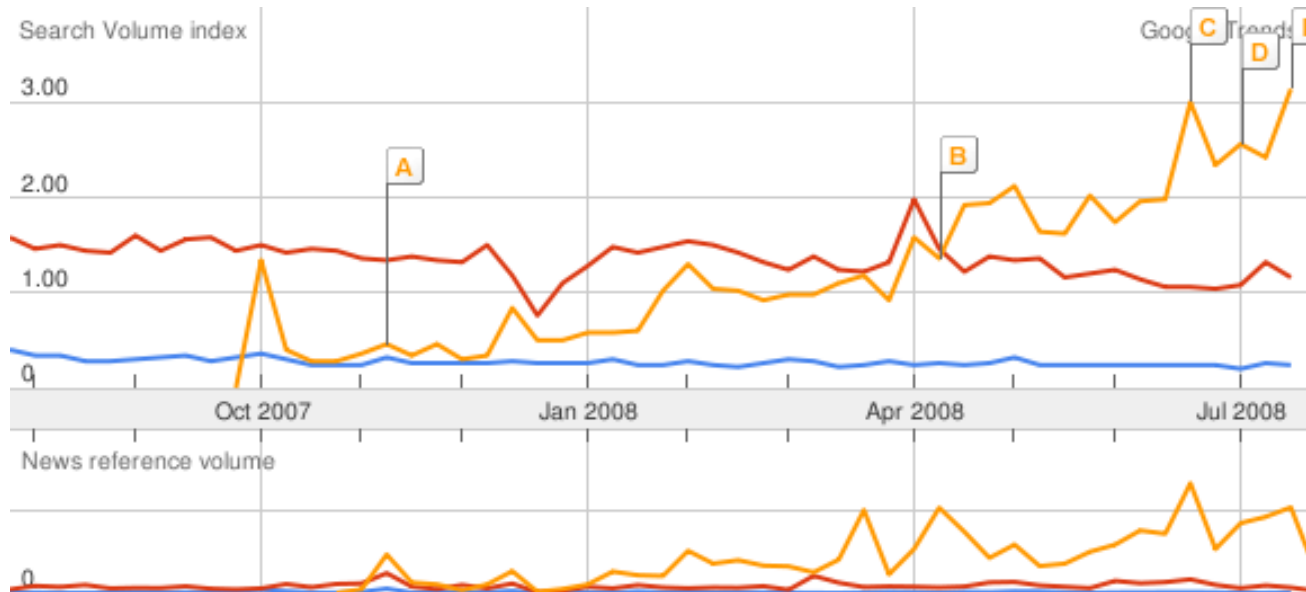
- 다양한 개방 시스템
- 확장성 및 표준에 집중
- Loosely coupled 시스템



SW 플랫폼 vs. 개방 플랫폼

독립 API를 가진 버추얼 머신	Platform	오픈 API를 가진 웹 서비스
독자적/폐쇄형	API	개방형/표준형
기술 장벽으로 진입이 어려움	Costs	빠르고 쉽게 개발 및 서비스 가능
기능 위주	Lock-in	데이터 위주
특정 플랫폼 소속 개발자	Developers	거의 모든 플랫폼 소속 개발자 포용
.Net Java	Examples	Google Amazon

The hype?



클러스터 컴퓨팅
클라우드 컴퓨팅
그리드 컴퓨팅



Definition

Five essential characteristics, which differentiate cloud computing from grid computing and other distributed computing paradigms:

- ***On-demand self-service.*** provision computing capabilities as needed automatically.
- ***Broad network access.*** available over the network and accessed through standard mechanisms.
- ***Resource pooling.*** computing resources are pooled with location independence
- ***Rapid elasticity.*** Capabilities can be rapidly and elastically provisioned.
- ***Measured Service.*** automatically control and optimize resource

클라우드 컴퓨팅 종류

IaaS

**Infrastructure
as a Service**

PaaS

**Platform as a
Service**

SaaS

**Software as
a Service**

- SaaS(Software as a Service)

- ☞ SW를 구매하거나 설치하지 않고 필요에 따라 인터넷을 통해 SW를 사용하는 서비스

- PaaS(Platform as a Service)

- ☞ 개발자가 SaaS를 개발/운용할 수 있도록 개발환경 플랫폼을 제공하는 서비스 (B2B)

- IaaS(Infrastructure as a Service)

- ☞ 서버, 스토리지 등의 IT인프라를 사용자가 필요한 만큼 제공하는 서비스. 기업의 전산시설을 위탁관리하는 IDC 같은 곳.

Service Model

IaaS

Infrastructure as a Service

- On-demand sharing physical infrastructures
- Users: System Administrator

 GOGRID

 amazon web services™

 flexiscale™

PaaS

Platform as a Service

- Platform for developing and delivering applications, abstracted from infrastructures
- Users: Developer

 Windows Azure

 Google App Engine

 force.com™
platform as a service

SaaS

Software as a Service

- Almost any IT services
- Users: End-user

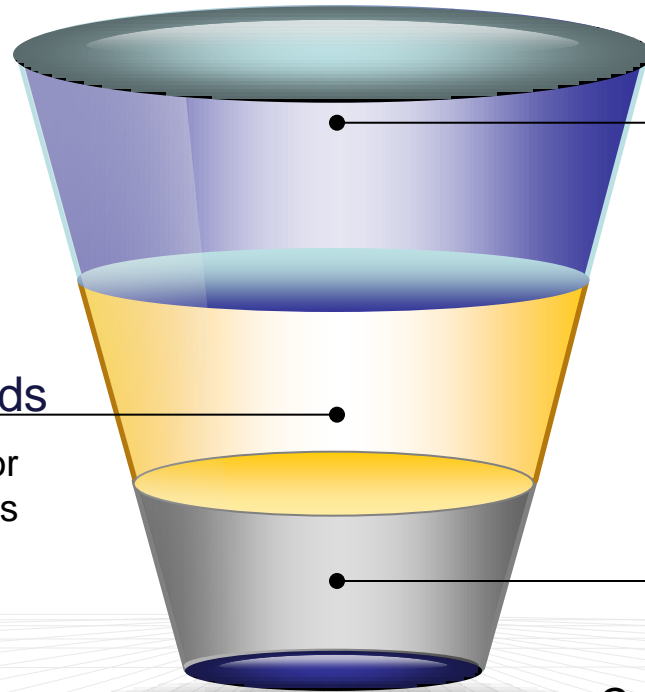

 skype™

WIKIPEDIA

 twitter™

 facebook

Type



Commercial Clouds



Private/Community Clouds

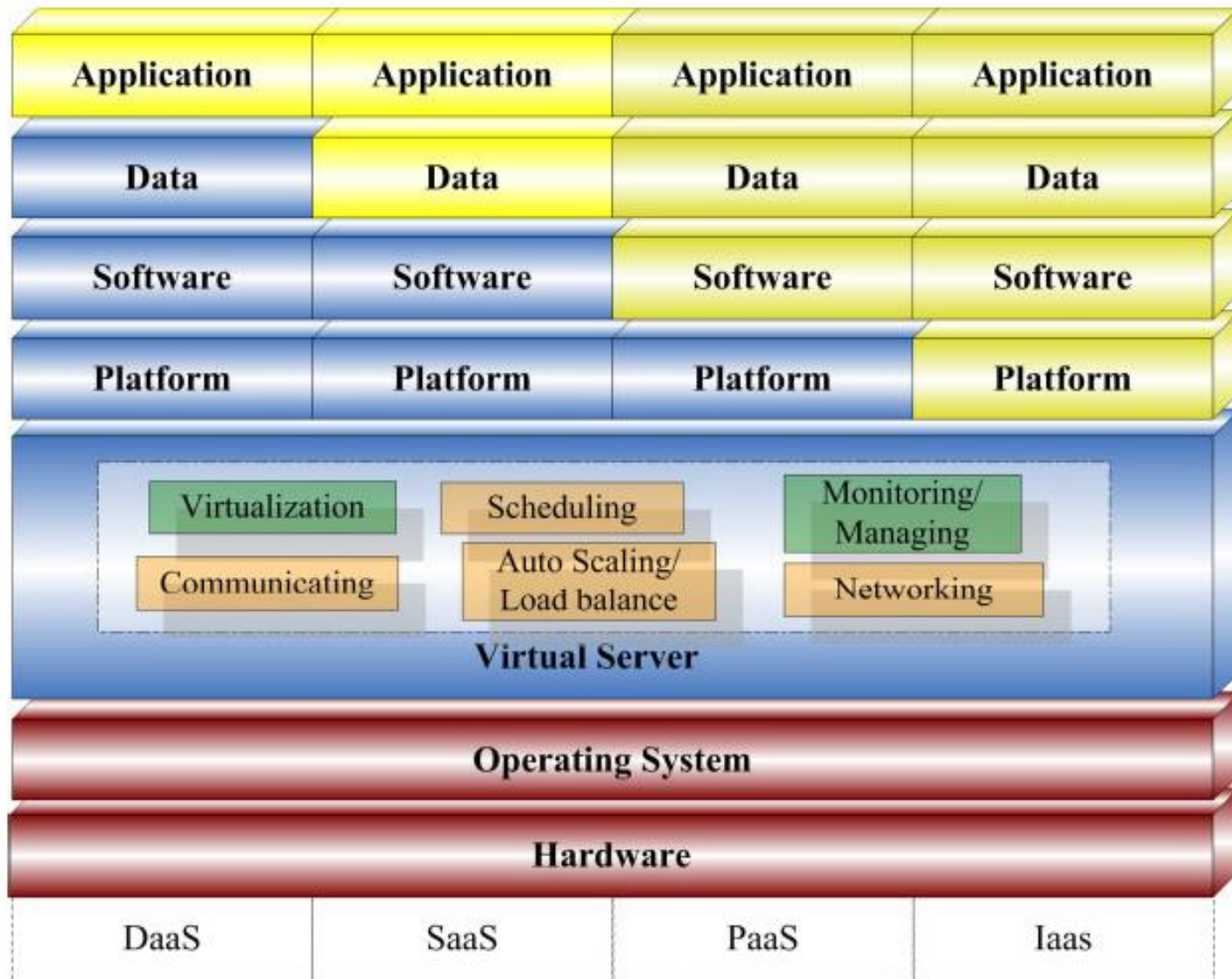
Build by commercial or open-source Solutions



Hybrid Clouds

Commercial clouds and private clouds: EC2 Vs Eucalyptus, EC2 Vs OpenNebular

Framework



Commons

Factors

- Pay per use
- Instant Scalability
- Security
- Reliability
- APIs

Advantages

- Lower cost of ownership
- Reduce infrastructure management responsibility
- Allow for unexpected resource loads
- Faster application rollout

Economics

- Multi-tenented
- Virtualisation lowers costs by increasing utilisation
- Economies of scale afforded by technology
- Automated update policy

Risks

Security Downtime

Dependency

Interoperability



AWS Stack

Distributed Infrastructure



Multi-AZ Services



S3



DynamoDB



RDS

Building Blocks



EC2 Instances



Elastic Load Balancer

Loosely Coupled Process Coordination



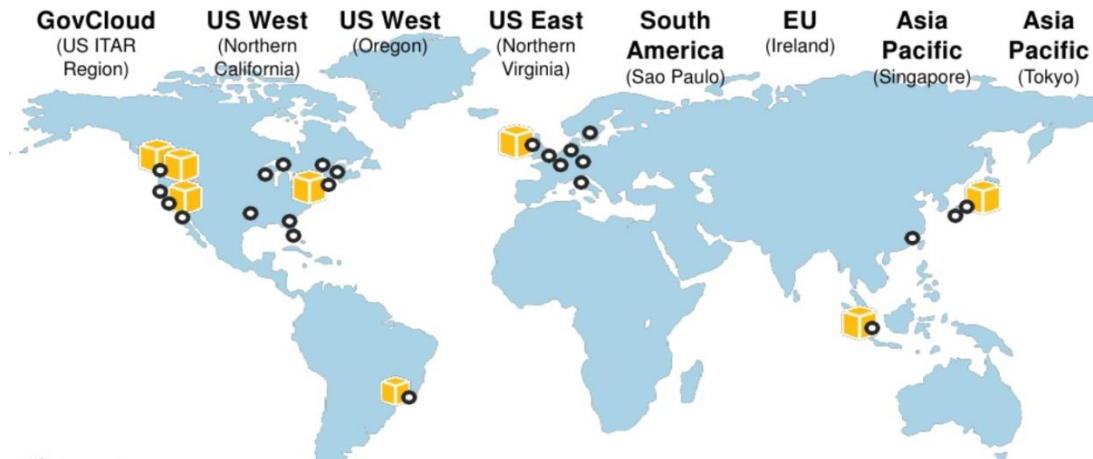
SWF



SNS



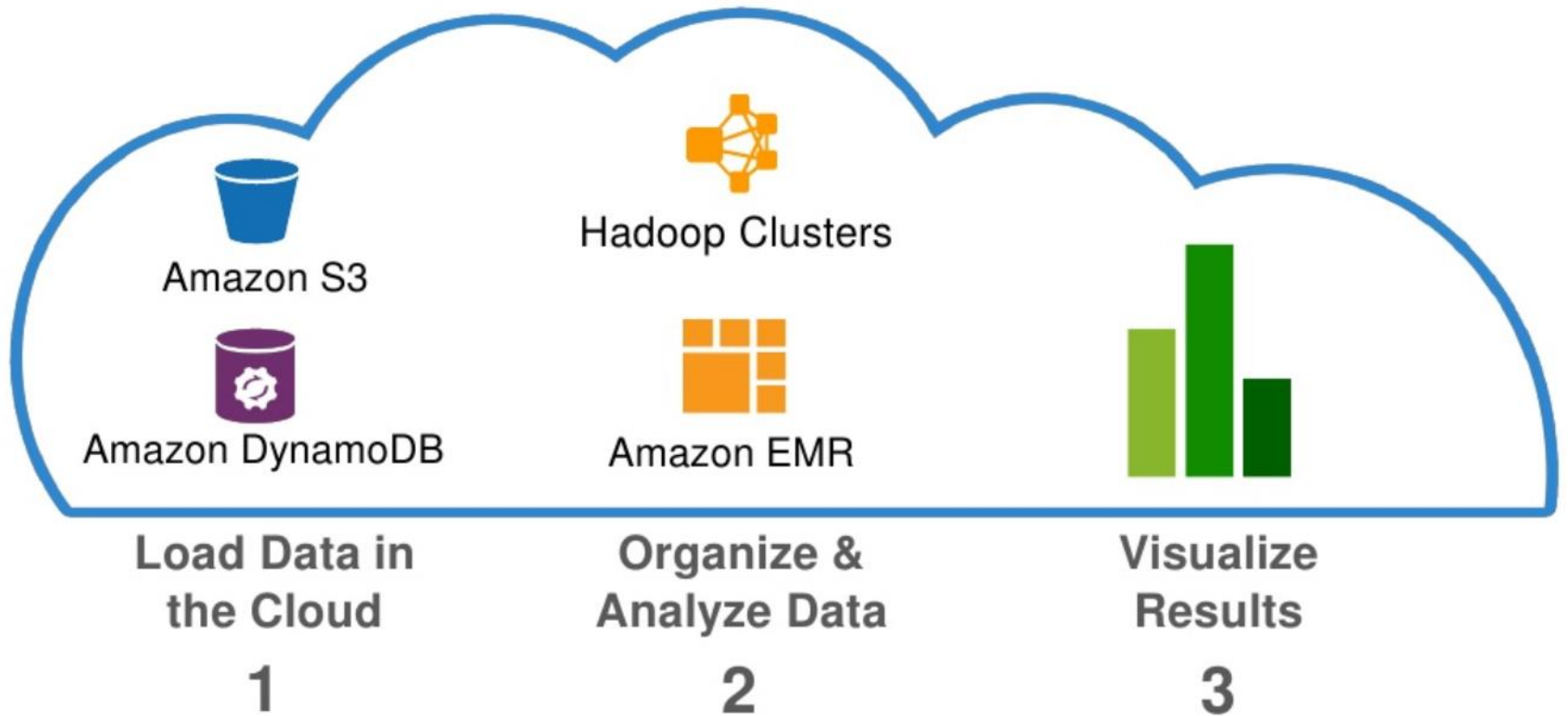
SQS



 AWS Regions

 AWS Edge Locations

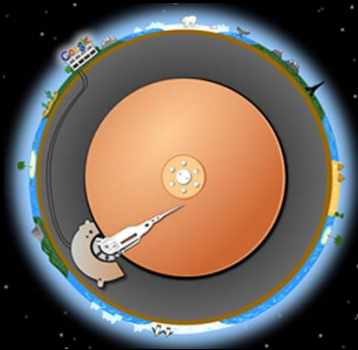
AWS Bigdata Analysis





Daum의 클라우드 컴퓨팅 활용 사례

1. 대용량 스토리지 제공



Gmail
by Google

100MB
per User

한메일넷

2004

TENTH
1GB



2006

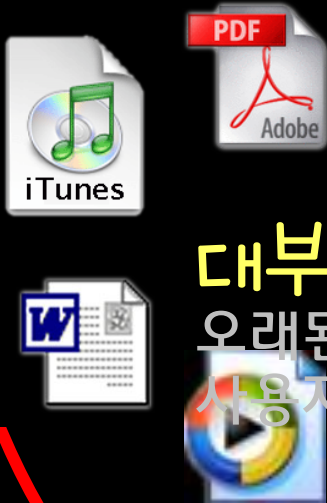
50GB



**명실 상부한
국내 최대
저장용량
제공 중**

2012

How to? Tenth₍₂₀₀₅₎ 프로젝트



대부분 첨부 파일은 30일이 지나면 안본다?
오래된 파일은 값싼 스토리지에 저장해 두고
사용자가 원할때 빨리 찾아주면 되지 않을까?

외부 분산 스토리지 기술

- Google GFS (2003)
- Yahoo HDFS (2007)
- NHN OwFS (2010)

30 days

Tenth vs. HDFS

- Tenth는 한메일, 카페 첨부 파일 등 대용량 파일을 저렴하게 저장하기 위한 분산 파일 시스템으로 2005년 부터 개발
 - 2006년 라이코스메일, 카페 도입
 - 2007년 한메일 기가 용량 도입
 - 2009년 동영상 업로드폼 도입
 - 2010년 다음 클라우드 도입
- 2012 현재, 파일 700억개(Dual Backup), 20페타바이트 (Zipped)

Tenth	비교	HDFS
2005	개발 시작	2006
C++	구현 언어	Java
첨부 파일을 저장하기 위해 하나의 스토리지 처럼 이용 가능	이용 목적	분산 시스템에서 파일 저장 용도로 활용
다중 (MySQL이용)	네임 노드	싱글
1~4MB (fixed chunks)	파일 형태	64MB (fixed blocks)
미지원	디렉토리 구조	지원함

2. 폭증하는 소셜 데이터



촛불집회 후 수 십만명이 동시에 글을 쓰면?

데이터베이스(RDBMS)는 과도한 쓰기(수정)를 하면 서버가 다운될 수 밖에 없다.



미디어다음 526 (월) | [홈](#) / [특수 18+](#) | [로그인](#) / [회원가입](#)

뉴스 | 스포츠 | **아고라** | 블로그뉴스 | 텔레비전 | 세계연 | 만화속세상

아고라 홈 | 토론 | 이야기 | 즐보드 | **청원** | 100분토론

청원 베스트 | **이슈청원** | [잠적청원](#) | [모금청원](#) | [컴퓨터인](#) | [추모서명](#) | [응원서명](#)

이슈 청원

서명 진행중 [1천만명서명]국회에 이명박 대통령 탄핵을요구합니다 [안단테](#)

현재 서명인원 **1,331,795** 명

서명목표 **10,000,000**명 | 마감일까지 220일 남았습니다

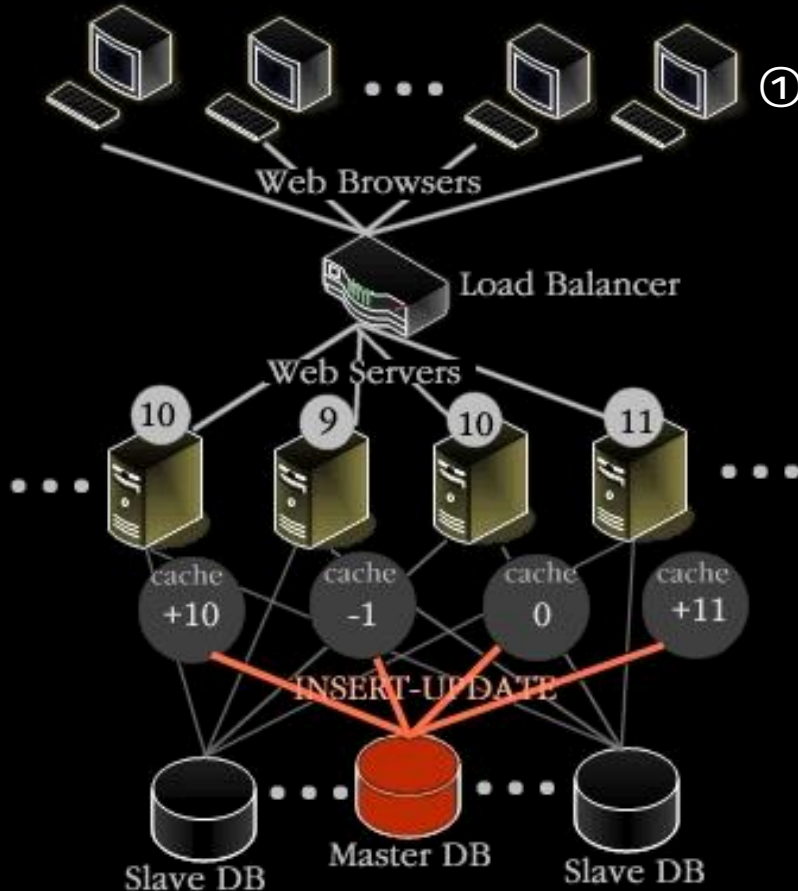
13% 달성

발의 08.04.06 | 마감 08.12.31

관련이미지 http://www.gobada.co.kr/csigsig_list.php

윤석찬(2008), 대통령 탄핵 청원으로 본 아고라의 기술적 도전
<http://blog.creation.net/260>

How to? 분산 데이터 캐시

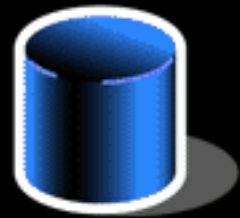


[일반적인 웹 서비스 분산 구조]

① 글 조회수를 서버 메모리에 모아 두었다가 어느 정도 모이면 한꺼번에 업데이트하면 어떨까?

최범균(2008), EHCACHE를 이용한 Gaia 캐시 구현,
<http://javacan.tistory.com/123>

② DB 앞에 데이터 캐시 서버를 두면 어떨까?
Pizza(2002) Santa(2008) Wcache(2010)



회사 내부에서 사용하는 인하우스 미들웨어로 JDBC 병목해결, 데이터 캐시 서버 및 대용량 캐시 서버 팜 등으로 진화.

3. 빅데이터 기술 활용 사례



- **로그 분석 사례**
 - 전사 로그를 통한 통계 분석
 - 광고 로그 분석을 통한 타겟팅
 - 검색 품질 랭킹 분석 및 개선
 - 광고 및 클릭 로그 분석을 통한 타겟팅
 - 카페 로그 분석을 통한 사용자 카페 추천
 - 게임 서버 로그 분석 등
- **데이터 분석 사례**
 - 쇼핑 하우 상품 클릭 분석 사례
 - 다음 Top 토픽 분석 및 추천 서비스
 - UCC 문서의 스팸 유저 필터링
 - 사물 검색 이미지 역색인
 - 자연어 처리 텍스트 분석
 - 모바일 광고 데이터별 매체 분석 등
- **연구 개발 사례**
 - 이미지 유사성 매칭 분석
 - 대용량 시맨틱 웹 검색 엔진 개발



- **서비스 적용** (MongoDB/카산드라)
 - 마이 아고라
 - 검색 광고 노출 최적화
 - 최근 방문 카페 저장
 - 사내 캐시 서버(Redis)
 - 사내 Git 저장소(Redis)
- **데이터 처리** (Hbase)
 - 검색 엔진 색인 문서 저장
 - 서버 모니터링 데이터 저장
 - 로그인 로그 저장
 - 카페 방문 로그 저장
- **서비스 분석**
 - 미디어 다음 실시간 분석
 - 모바일/PC탐 실시간 분석
- **데이터 수집**
 - Twitter 실시간 데이터 수집기



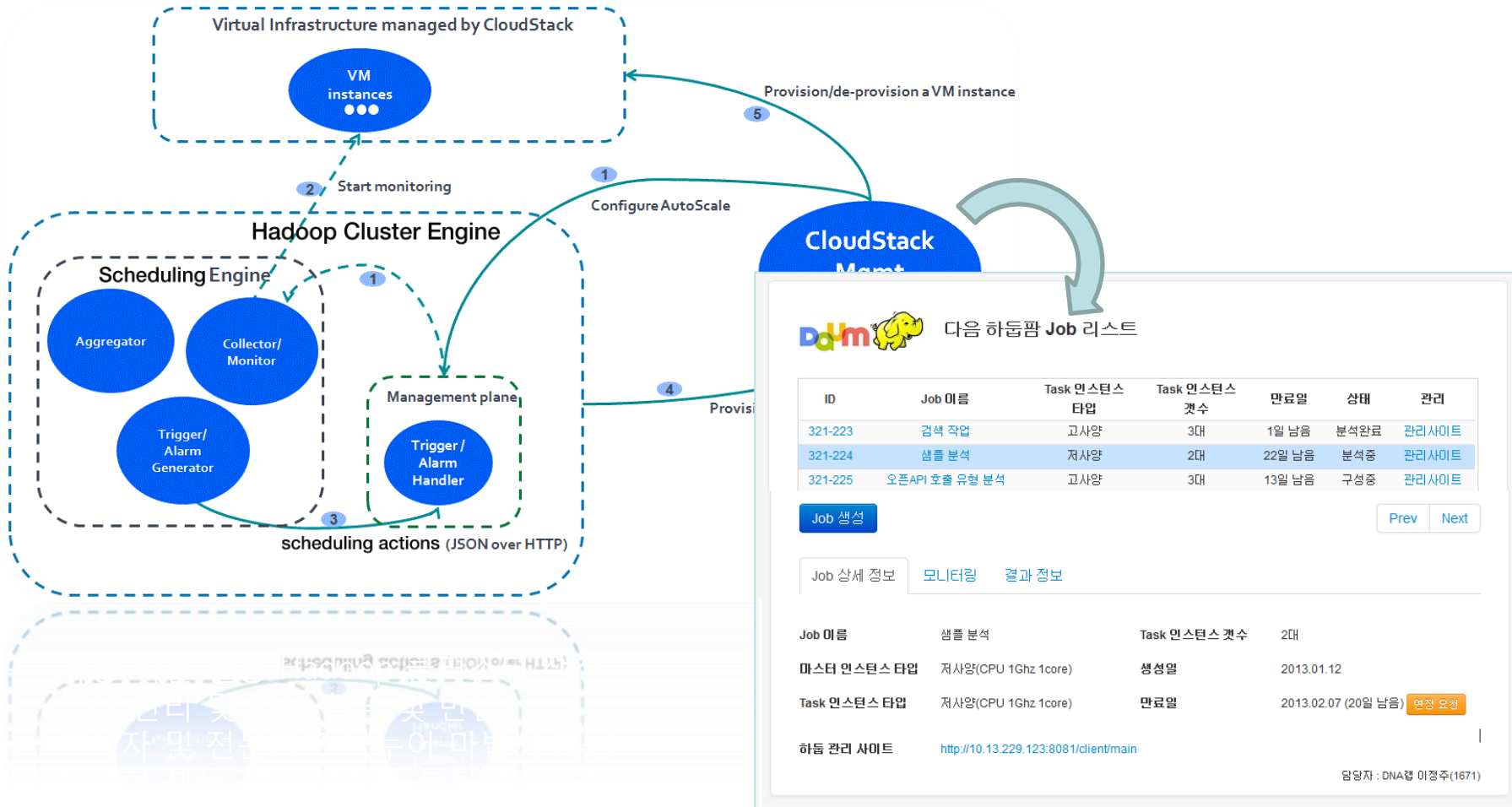
가상 서버 테스트베드

- 서비스용 테스트 서버
 - VM웨어 기반 시스템
 - 테스트 서버, 개발 서버
 - 총 2,000 가상 서버
- 개발자용 테스트 서버
 - 클라우드스택 기반
 - 개인 실험용
 - 총 600 VM



사내 개발자를 위한 Hadoop Farm

- 오픈 소스 CloudStack과 Hadoop 클러스터를 통한 유연한 분석용 온디멘드 작업 서비스 프로토타입 및 제공



기타 클라우드 서비스

- Amazon Web Services
 - 클라우드 플랫폼 개발 환경, 테스트베드
 - NIS (Next Incubation Studio)
- Cloud Foundry
 - 사내 개발 및 실험 환경 (Sandbox)
- Redis Cache Cloud
 - 미디어다음, 카페, 블로그, Tistory 등
- 애조로 DNS
 - 사내 Dynamic DNS 서비스
 - 개인 개발 환경 및 각종 백엔드 호스트 연결



Q&A

더 자세한 실습 및
사례 데모는 워크샵에서...

Twitter: @channyun

E-mail: channy@daumcorp.com