

# BIG DATA: TECHNOLOGIES AND APPLICATIONS

C15 Cloud Computing and Summary

Il-Yeol Song, Ph.D.  
College of Computing & Informatics  
Drexel University  
Philadelphia, PA 19104

## In this Lecture

- Cloud computing
- Summary of the course



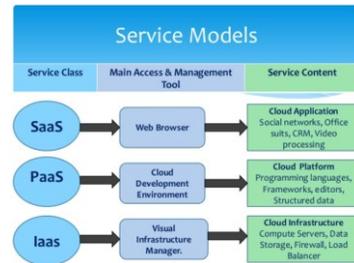
## Cloud Computing

- Cloud Computing is the delivery of computing as a **service** through a Web browser using **"pay as you use and need"**:
  - Including servers, storage, software, and applications
- Computing as a utility: **"access anytime and anywhere"** via Web
- Major Benefits
  - Reduce costs for IT infrastructure /maintenance
  - Quick engagement  
E.g., Use AWS for a Hadoop framework

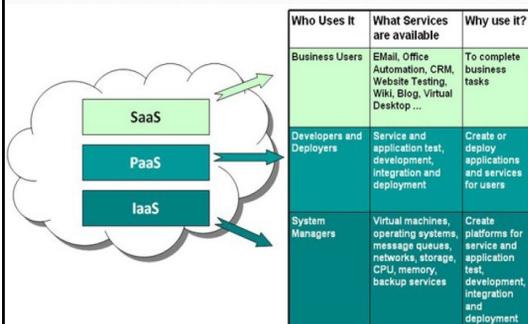


## The Three Cloud Services

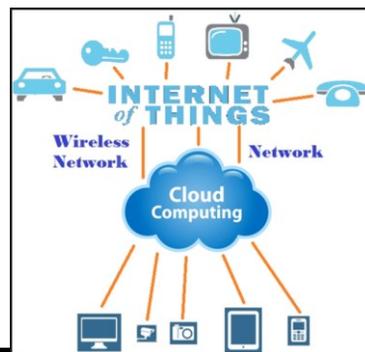
- SaaS: Software as a service
- PaaS: Platform as a service
- IaaS: Infrastructure as a service



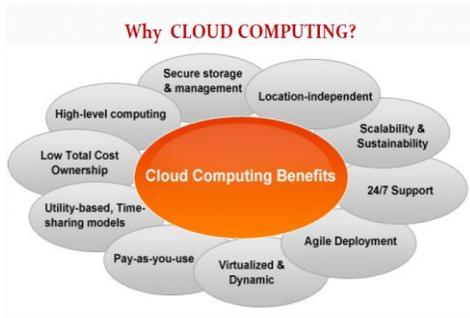
## The Three Cloud Services



## Cloud and IOT



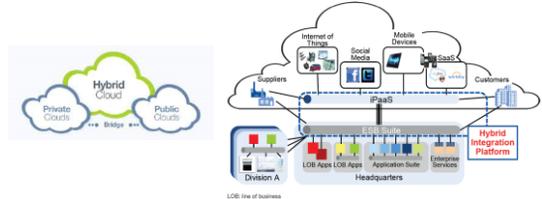
## Why cloud?



Il-Yeol Seog, Ph.D.

## Hybrid Cloud Computing

- Store *private/sensitive/critical data* in *on-premise* and *sharable data* in a **public cloud**
- Valuable for **dynamic** or **highly changeable** workloads.



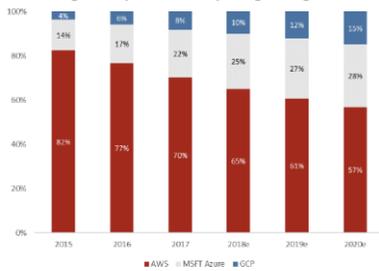
Source: <http://searchcloudcomputing.techtarget.com/definition/hybrid-cloud>  
Source: <http://eng.itsm.com/hybrid-cloud-look-out-to-the-next-big-thing/>

Il-Yeol Seog, Ph.D.

8

## Cloud Market Share

Market share among the major Clouds – expecting share gains from GCP



Il-Yeol Seog, Ph.D.

## Cloud Computing

- **Concerns:**
  - Performance is dependent on others
  - Reliability
    - What if the remote server is down?
    - Cloud vendors will do better than your organization
  - Control of data
    - Who owns the data when your data is stored in a foreign cloud? Which data-governing policy?
    - Government intrusion
    - Legal liability

Il-Yeol Seog, Ph.D.

## Cloud Computing

- **Concerns:**
  - Cost of data migration, loading, and integration
    - Unexpectedly high cost –eventually
    - Each cloud is another silo
  - Security:
    - Someone is looking at your data!
    - Cyber attacks
    - Insider threats
  - Privacy
    - What if confidential data is leaked?

Il-Yeol Seog, Ph.D.

## Cloud Computing



- Cloud is a strategy
  - Provides worry-free maintenance; rapid scalability; incremental cost as you pay as you use
  - Provides system and data backup, archival, and recovery
  - Cloud itself will not solve data integration and management issues.
  - Just introduces efficiency to the process
  - Hence, cloud needs good data governance, high quality metadata, and well-understood data integration process

Il-Yeol Seog, Ph.D.

| 12

## So What?

- As digital footprints become larger, more will be cloud-based, convenient for access and the security of data redundancy.
- Security & Privacy in cloud computing are major concerns.
  - Data in Cloud should be stored in encrypted form.
- Cloud is a tradeoff between cost and security & privacy.
- The cloud is a strategy.

## Required readings

[What is cloud computing?](#)  
[Security issues associated with the cloud.](#)  
[Digital footprints and privacy.](#)

## Creating Big Data Projects

1. Add more detailed transaction data
2. Add unstructured data such as Web, social media, blog, web log data,
3. Add dark data
4. Add low-latency, real-time, data
5. Integrate predictive analytics

## A Summary of BD Technologies (I)

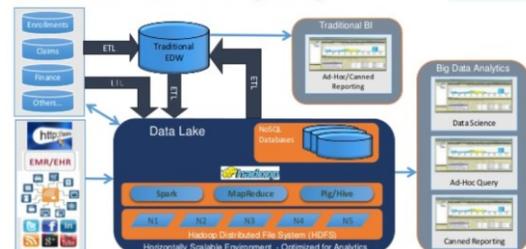
- Big data technologies address issues on Web, Mobile, Social, Cloud, Big Data Analytics, and AI
- Big Data can be characterized by 5Vs
- Hadoop is still good for large unstructured batch applications
- Spark will dominate real-time or interactive applications
- NoSQL systems were designed to deal with big data up to petabyte level
  - RDBMSs are good for multi-gigabyte level applications
  - Most NoSQL systems use Schema-on-Read and relaxes the ACID property
  - NewSQL systems are growing

## A Summary of BD Technologies (II)

- In-memory computing is widely used in NoSQL, NewSQL, and analytics
- Hadoop, EDW, and Data Lake will co-exist
  - ✓ DW contains refined data, while a data lake stores raw data
  - ✓ Data lakes are implemented using Hadoop or in Cloud
- Driving force behind IR 4.0: Big Data, AI, and IOT
- AI makes the system intelligent and smart
- Machine learning platforms are increasingly automated and could support most phases of analytical lifecycles.
- Cloud will become more popular
- Watch out blockchain technology and its applications!

## Integrated Modern Big Data Architecture

### The Evolution of Modern Data Engineering



## A Summary of BD Applications (I)

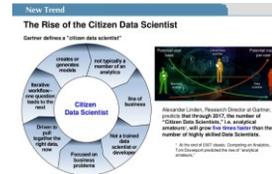
- Big Data enables new things
- Identify business opportunities through datafication
- For big data projects, apply data analytic life cycles
- Begin from market needs or user needs and then consider technologies
- Asking right business questions are important
- Ability to translate business questions into data science problems are critical
- Big data driving force: AI applications, IoT, Smart City, Smart Health, smart farming, smart manufacturing.
- Predictive analytics solutions analyse patterns found in big data, create a model that can predict potential future outcomes.
- Deep learning is widely applied; good for large scale noisy data

## A Summary of BD Applications (II)

- Create Big Data Projects with the 5 ideas
- Manage big data projects with the data analytic lifecycle in mind.
- Value can be obtained when sticking to analytics life cycle and following best practice
- Cloud-based analytics are getting automated and powerful

## A Summary of BD Applications (II)

- Citizen data scientists will be in demand
  - Become a T-shaped data scientist!



- Develop your career to become a CDO!



## A Summary of BD Applications (II)

- Consider new AI applications
  - AI is the new electricity
  - AI Speaker, Digital twins, Virtual Assistants, Apple Siri, chatbots, Watson, AlphaGo, AI Lawyer, AI Stock trader, Driverless cars, Robots
  - AI application to business: recommender sys, marketing, finance, insurance, healthcare, agriculture, manufacturing
  - New AI applications in public services, government, education
  - Think about new applications that utilize AI technology
    - Uber: Google Map + Apply pay + Software



## Predicting the BD Future?

*The best way to predict the future is to invent it!*

--- Alan Kay

감사합니다 (Korean)    多謝 (Traditional Chinese)    ටොන්කුම (Sinhala)  
 Спасибо (Russian)    Gracias (Spanish)  
 Thank You (English)  
 شكراً (Arabic)    Obrigado (Brazilian Portuguese)  
 Grazie (Italian)    多謝 (Simplified Chinese)    Danke (German)  
 நன்றி (Tamil)    ありがとうございます (Japanese)    धन्यवाद (Hindi)