

# TK1 - 4th Theory Exercise

Shule Liu, Kecen Li, Zhen Chen, Letian Feng, Xin Zhang

30. November 2016

## Task 1

Cloud computing provides hosting options for data and services such that users can access them via the Internet. With highly virtualized infrastructure, its high scalability implies that cloud computing is capable of handling the elastic demands from service providers and users. Cloud computing users are provided with higher computing capability comparing to their local machines and are guaranteed with higher efficiency and dependability. Since the technical infrastructure is maintained and monitored by the cloud systems by professionals, it makes the optimization of resource usage feasible and effective and users can focus on their services over the cloud.

A good example of cloud computing is Amazon EC2 service.

## Task 2

Public Cloud, Private Cloud, Community Cloud, Hybrid Cloud

## Task 3

Infrastructure as a Service

Computing infrastructures such as processors, memories, storage and bandwidth are virtualized and provided from a network of servers by the cloud system. Subscription of computing infrastructures is highly customisable by user based on their individual needs and they pay accordingly as well. Above the infrastructural layer, user has to manage service and probably operating system on their own. Example, Amazon EC2.

Platform as a Service

Instead of providing virtualized computing infrastructures directly to the user, PaaS builds upon the infrastructure with an additional layer that involves operating system, software framework etc. (referred as the platform) to facilitate user requirements such as web hosting and application development. Example, Google App Engine.

Software as a Service

SaaS providers manage a full stack of hardware, middleware and software and provides services to users that are reachable from anywhere through internet. User can access the service without any instalment or other prerequisite through a browser in most of the cases. Example, Google Apps.

## Task 4

Chunk server in GFS functions as the primary storage facilities to store data chunk of size 64 MB by default.

During read operation, chunk server return requested data back to clients. During write operations, primary chunk server forwards request to secondary chunk servers. Secondary chunk servers reply primary chunk server upon completion of the request and primary chunk server reply to the client when all replicas are updated.

## Task 5

Relational database retrains a well defined data model for each table and every entry in a table comply to the table schema while in key-value pair store, entries are simply dumped into tables without necessarily being compatible to a certain schema. Also, it lacks a data model that defines the relationship between entries.

RDBMS normalised its data to avoid duplicated while key-value store doesn't do this.

RDBMS uses SQL for data manipulation and offers functions for aggregation and complex filtering while in key-value store, data is manipulated through basic APIs and only simplistic filtering is supported since processing is not a primary concern.

RDBMS data schema is often mapped to application data through Object Relational Mapping (e.g. Entity Framework on .Net platform) while in key-value store, application data is flattened into key-value pairs directly and there's no effective way to map key-value pairs back to an data object.

Prefer key-value pair for binary file storage.