introduction to cloud computing

cloud computing

- pay per use, on-demand web-based access, shared pool of computing resources
- IaaS, PaaS, SaaS

IaaS

virtual machines on shared hardware, virtual networks, storage

public cloud / private cloud (companies)

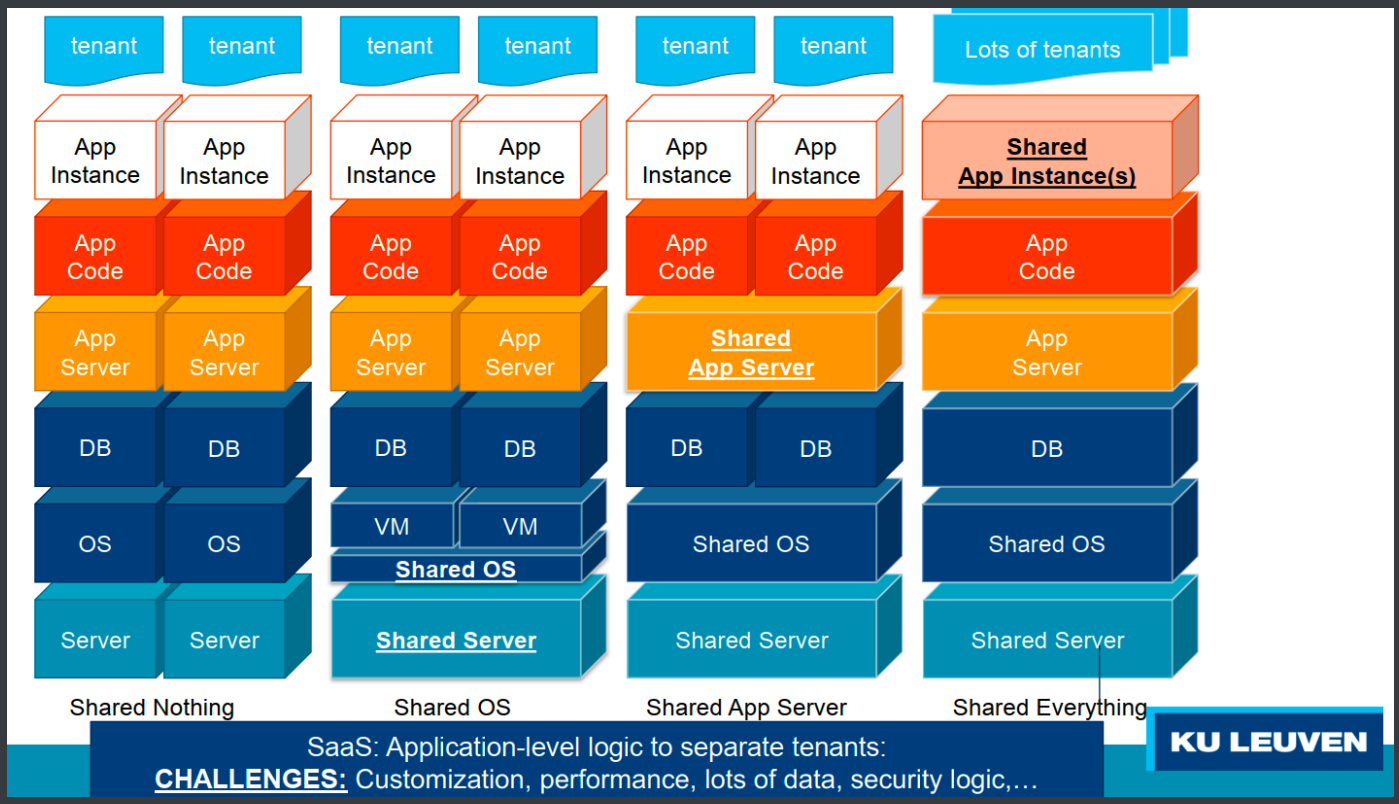
PaaS

- middleware as a service, e.g. google AppEngine, MS Azure
- scalability first
- resource sharing (impact on performance / failure isolation)
- serverless model

  devops (deployment) is quite annoying ->

  let cloud provider manage automatic scaling & sw stack

SaaS



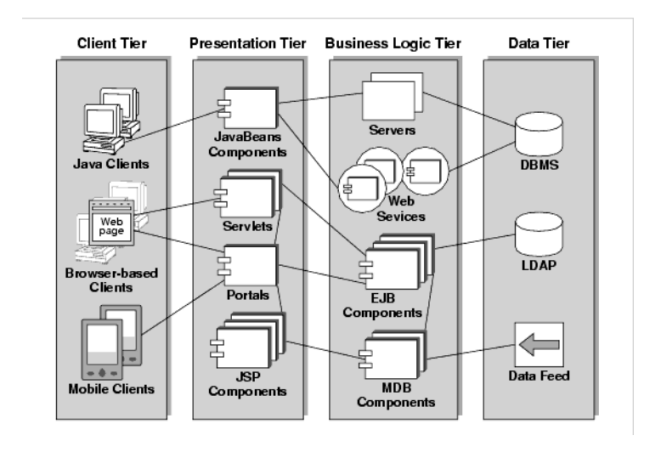
multi-tier architecture

2-tier = client + server (data+app logic)

3-tier = user interface (presentation) + data storage + business logic

multi-tier = client + presentation + application + data

- client
- presentation (server-side): 展示其他层的运算结果
- application tier: perform detailed processing
- data: data persistence mechanisms, provides API of managing stored data to the application tier



which tier is in "the" cloud?

client ❌, presentation maybe, application ✅, data ✅

core characteristics of cloud

1. higher degree of distribution

   different data centers/locations, replication

   availability & performance
2. multi-tenancy

   shared resources, multiplexing
3. elasticity

   dynamically provisioning, scaling horizontally, flexibility
4. delivery as a service

   client has no upfront investments in infrastructure, pay per use
5. self-service

   scalability, rapidly respond

case study: google AppEngine

case study: Microsoft Azure