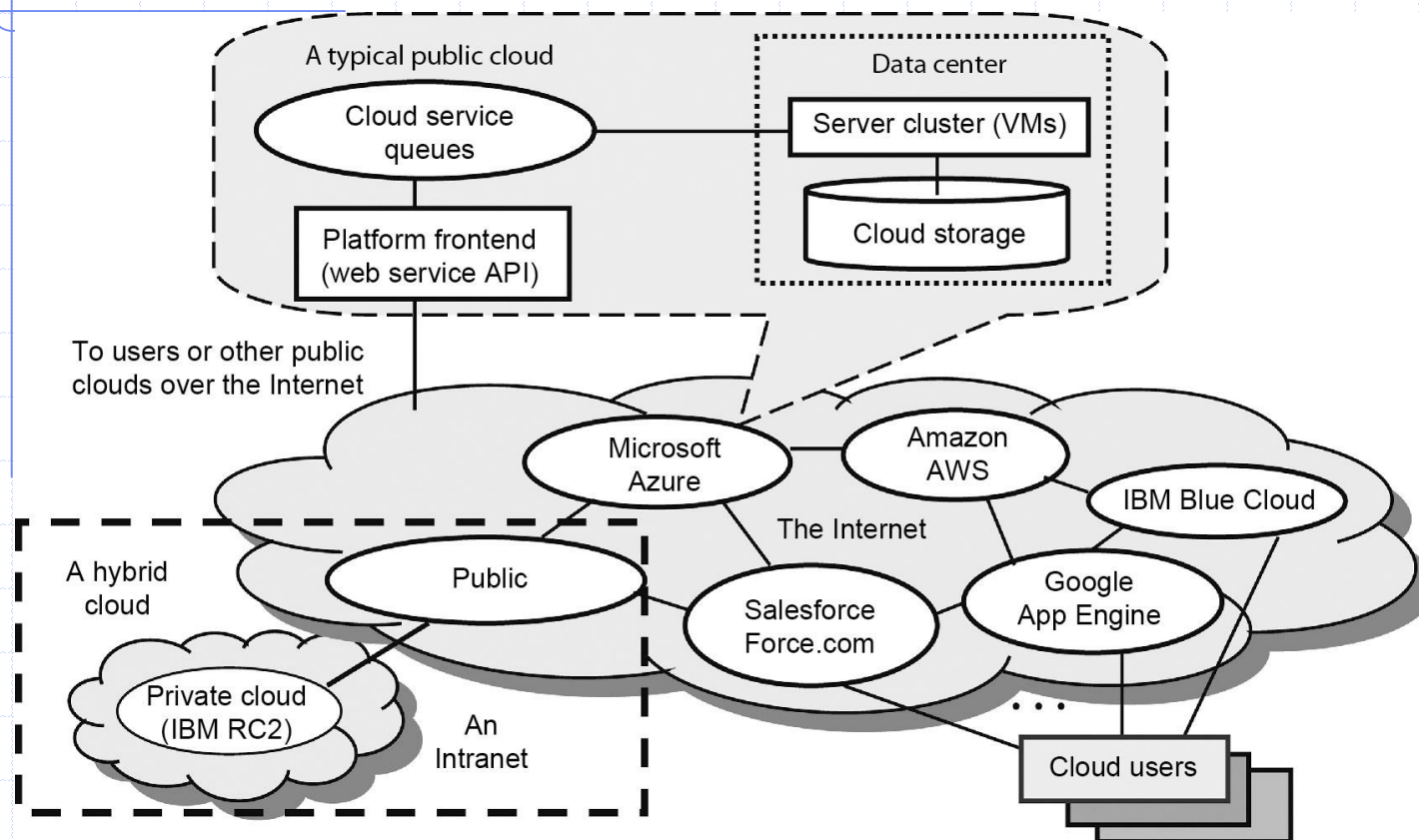


Arhitektura platforme oblaka i virtuelizovani centri za podatke

- ❖ Oblak

- ❖ Virtuelizovani centri za podatke

Vrste oblaka

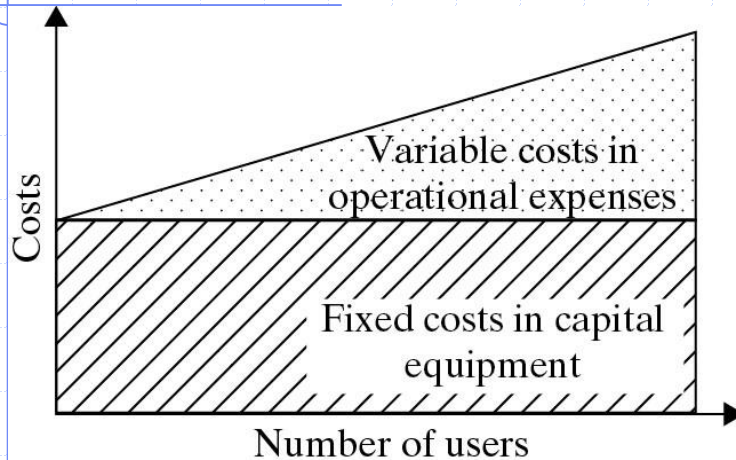


- ◆ Javni oblaci (Azure, AWS, Blue Cloud App Engine, Force.com)
- ◆ Privatni oblaci (RC2)
- ◆ Hibridni oblaci

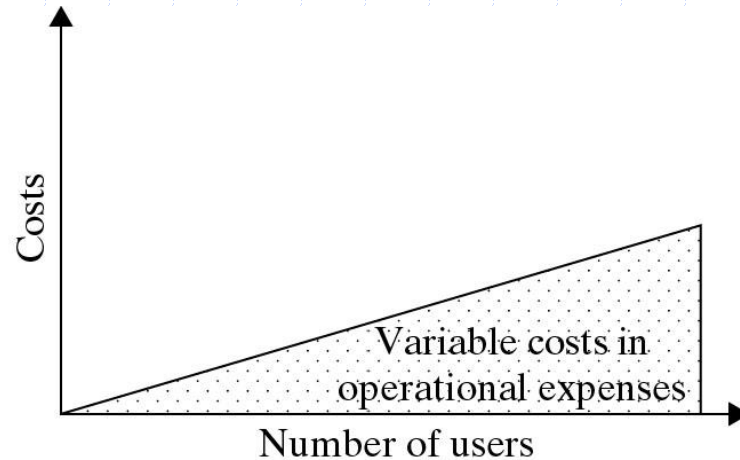
Javni oblaci naspram privatnih oblaka

Karakteristike	Javni oblaci	Privatni oblaci
Vlasništvo tehnologije	Vlasništvo dobavljača usluge	Vlasništvo individualne organizacije
Rukovanje sa obezbeđenim resursima	Stvaranje i rukovanje VM instancama unutar vlasničke infrastrukture; promoviše standardizaciju, čuva kapitalne investicije, obezbeđuje aplikacionu fleksibilnost	Rukuje klijent; podržava kastomizaciju i nudi veću efikasnost
Metode distribucije opterećenja i politike vođenja	Rukovanje opterećenjem bez komunikac. zavisnosti; distribucija podataka i VM resursa; rasterećuje se prelivno opterećenje	Rukovanje opterećenjem dinamički, ali može bolje uravnotežiti opterećenje; distribucija pod. i resursa VM
Zaštita i obezbeđivanje privatnost pod.	Javno dostupni kroz udaljenu spregu	Pristup je ograničen; obezbeđeno testiranje pre upotrebe i nametnuta privatnost pod. i politike zaštite
Primeri platformi	Google App Engine, Amazon AWS, Microsoft Azure	IBM RC2

Isplativost Računanja u oblaku spram iskorišćenja Centra za pod.



(a) Traditional IT cost model



(b) Cloud computing cost model

- ◆ U tradicionalnom IT modelu troškova postoje i fiksni troškovi
- ◆ U modelu troškova za oblak postoje samo varijabini troškovi
- ◆ Profit sa oblakom je veći nego sa tradicionalnim centrom za pod.

$$\text{RadniSati}_{\text{oblak}} \times (\text{prihod} - \text{trošak}_{\text{oblak}}) \geq \text{RadniSati}_{\text{centar}} \times (\text{prihod} - \text{trošak}_{\text{centar}} / \text{iskorišćenje})$$

Korisnici oblaka i nivoi rukovanja

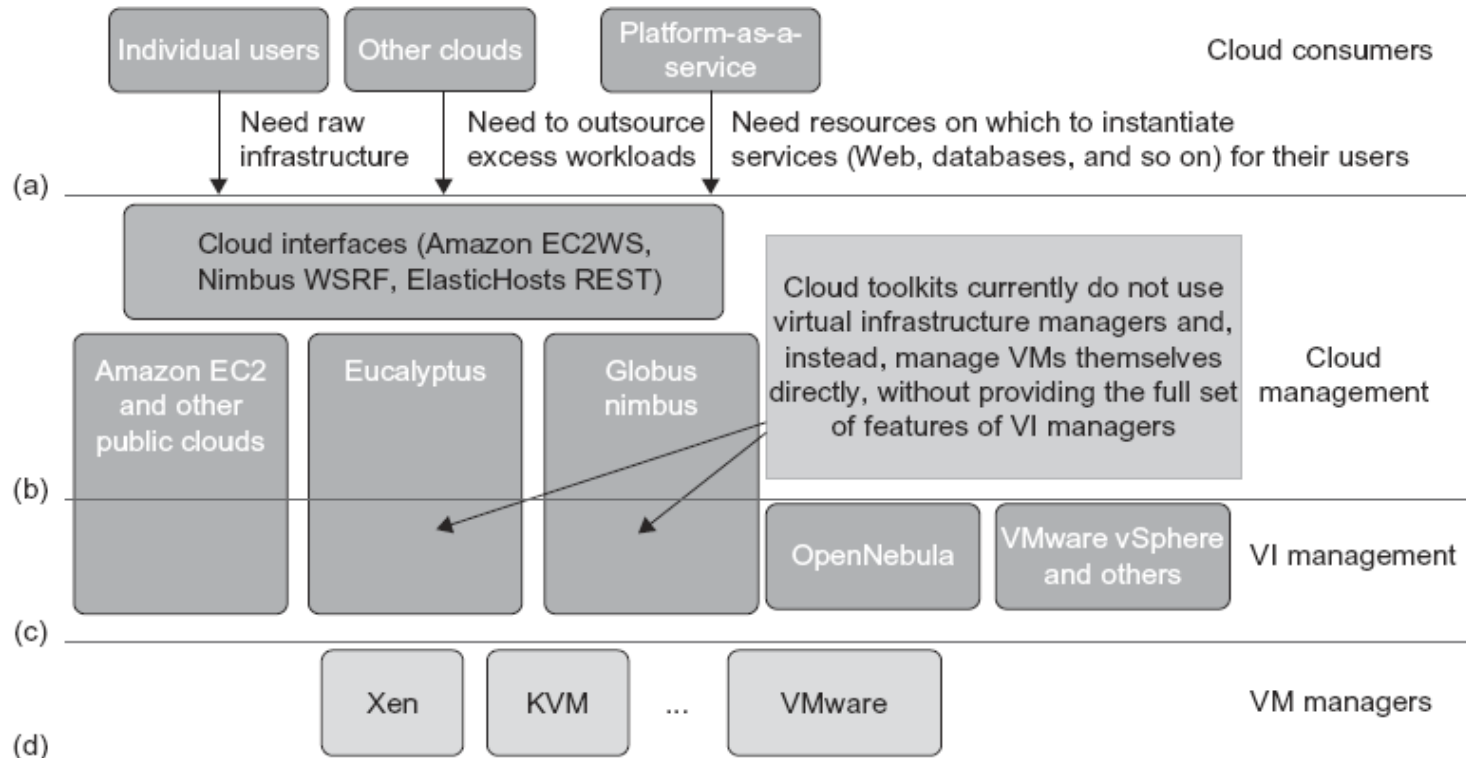


FIGURE 4.4

Cloud ecosystem for building private clouds: (a) Consumers demand a flexible platform; (b) Cloud manager provides virtualized resources over an IaaS platform; (c) VI manager allocates VMs; (d) VM managers handle VMs installed on servers.

(Courtesy of Sotomayor, et al. [68])

- ◆ Korisnici oblaka: pojedinci, drugi oblaci, drugi PaaS servisi
- ◆ Nivoi rukovanja: ruk. oblakom, ruk. VI, ruk. VM

Modeli i nivoi servisa oblaka

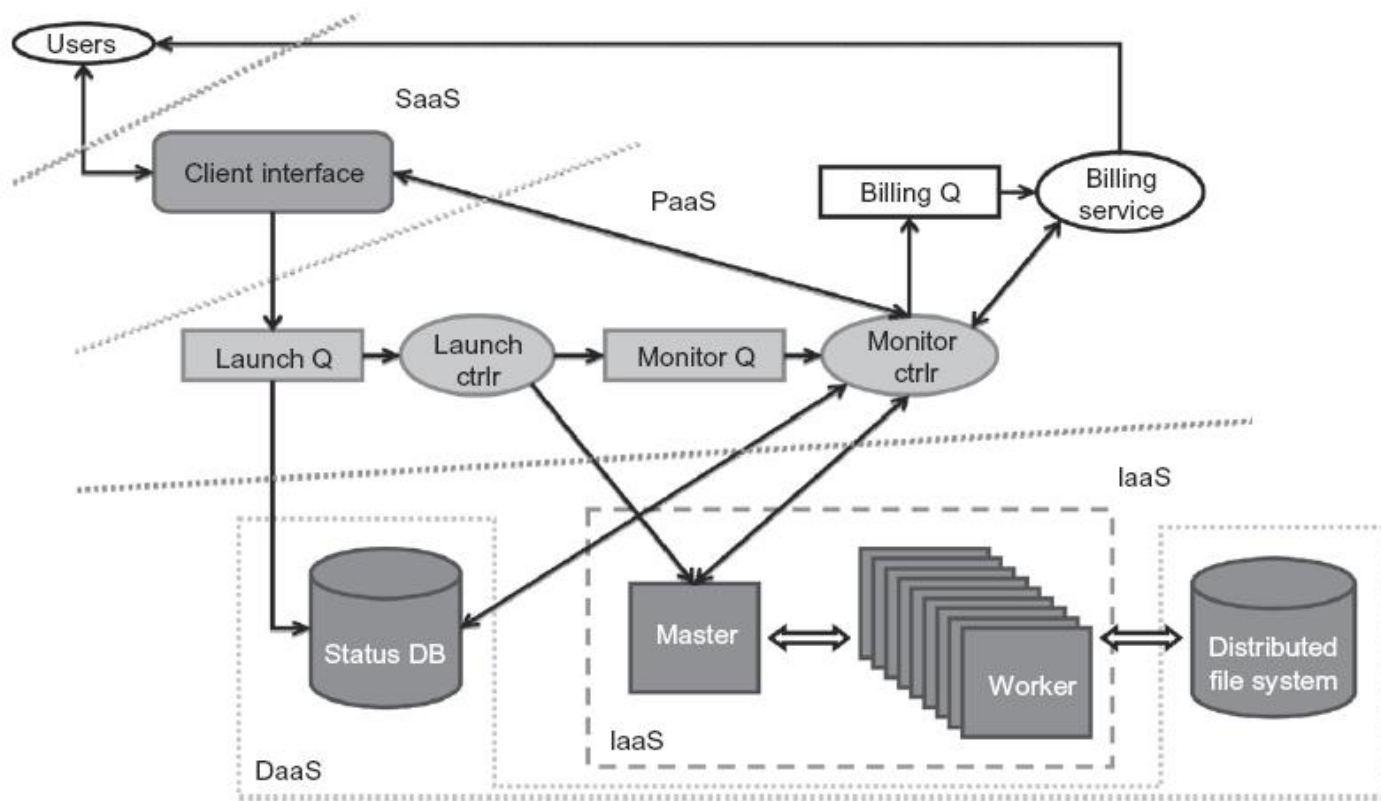


FIGURE 4.5

The IaaS, PaaS, and SaaS cloud service models at different service levels.

- ◆ Modeli servisa od gore na dole: SaaS, PaaS, IaaS
- ◆ SaaS nudi klijentski interfejs
- ◆ PaaS redovi za lansiranje, nadzor i naplatu

Infrastruktura kao usluga (IaaS)

◆ IaaS je osnovni model usluge oblaka

◆ Vlasnici:

- Nude računare, kao fizičke ili kao VM, i druge resurse
- VM rade kao gosti hipervizora, kao što je Xen ili KVM
- Naplaćuju IaaS usluge na osnovu količine dodeljenih i potrošenih resursa

◆ Korisnici:

- Instaliraju svoje slike OS i svoj aplikacioni softver

◆ Primeri:

- Amazon EC2, Rackspace Cloud, Terremark, i Google Compute Engine

Platforma kao usluga (PaaS)

◆ Vlasnici:

- Nude računarsku platformu koja obično uključuje OS, izvršno okruženje programskog jezika, bazu podataka, i mrežni (web) server

◆ Korisnici:

- Razvijaju aplikacije i izvršavaju ih na platformi oblaka bez troškova i poteškoća kupovine i održavanja pozadinskih slojeva HW i SW

◆ Primeri:

- Amazon Elastic Beanstalk, Cloud Foundry, Heroku, Force.com, EngineYard, Mendix, Google App Engine, Microsoft Azure i OrangeScape

Softver kao usluga (SaaS)

◆ Vlasnici i Korisnici:

- Vlasnici upravljaju aplikacionim softverom u oblaku, a korisnici oblaka pristupaju softveru iz klijenata oblaka

◆ Model cene:

- Najčešće je jednokratna predplata po korisniku, pa je cena skalabilna i podesiva ako se korisnici dodaju ili uklanjaju u bilo kojoj tački

◆ Primeri:

- Google Apps, innkeypos, Quickbooks Online, Limelight Video Platform, Salesforce.com, i Microsoft Office 365

Računar skale skladišta (Warehouse-Scale Comp, WSC)

- ◆ Obezbeđuje internet usluge
 - Pretrage, socijalne mreže, mape, deljenje videa, kupovine, email, računanje u oblaku, itd.
- ◆ Razlike u odnosu na HPC klastere:
 - Klasteri imaju procesore i mrežu više performanse
 - Klasteri naglašavaju paralelizam na nivou niti, WSCs naglašava paralelizam na nivou zahteva
- ◆ Razlike u odnosu na tradicionalne centre za pod:
 - Centri konsoliduju mašine i SW na jednoj lokaciji
 - Centri za podatke naglašavaju VM i heterogenost HW u cilju posluživanja raznolikih klijenata

Zahtevi za projekat WSC (1/2)

◆ Cena-performansa

- Male uštede se sabiraju

◆ Energetska efikasnost

- Utiče na distribuciju napajanja i hlađenje
- Rad po džulu

◆ Pouzdanost na osnovu redundantnosti

◆ Mrežni U-I

◆ Obrada interaktivnih i paketskih radnih opterećenja

Zahtevi za projekat WSC (2/2)

- ◆ Obiman računski paralelizam nije važan
 - Većina poslova su potpuno nezavisni
 - Paralelizam na nivou zahteva
- ◆ Tekući troškovi su važni
 - Potrošnja el energije je primarno ograničenje za projektovanje sistema
- ◆ Skala i njene mogućnosti i problemi
 - Može se priuštiti namenski sistem jer WSC zahteva veću nabavku

Tipičan tlocrt Centra za podatke

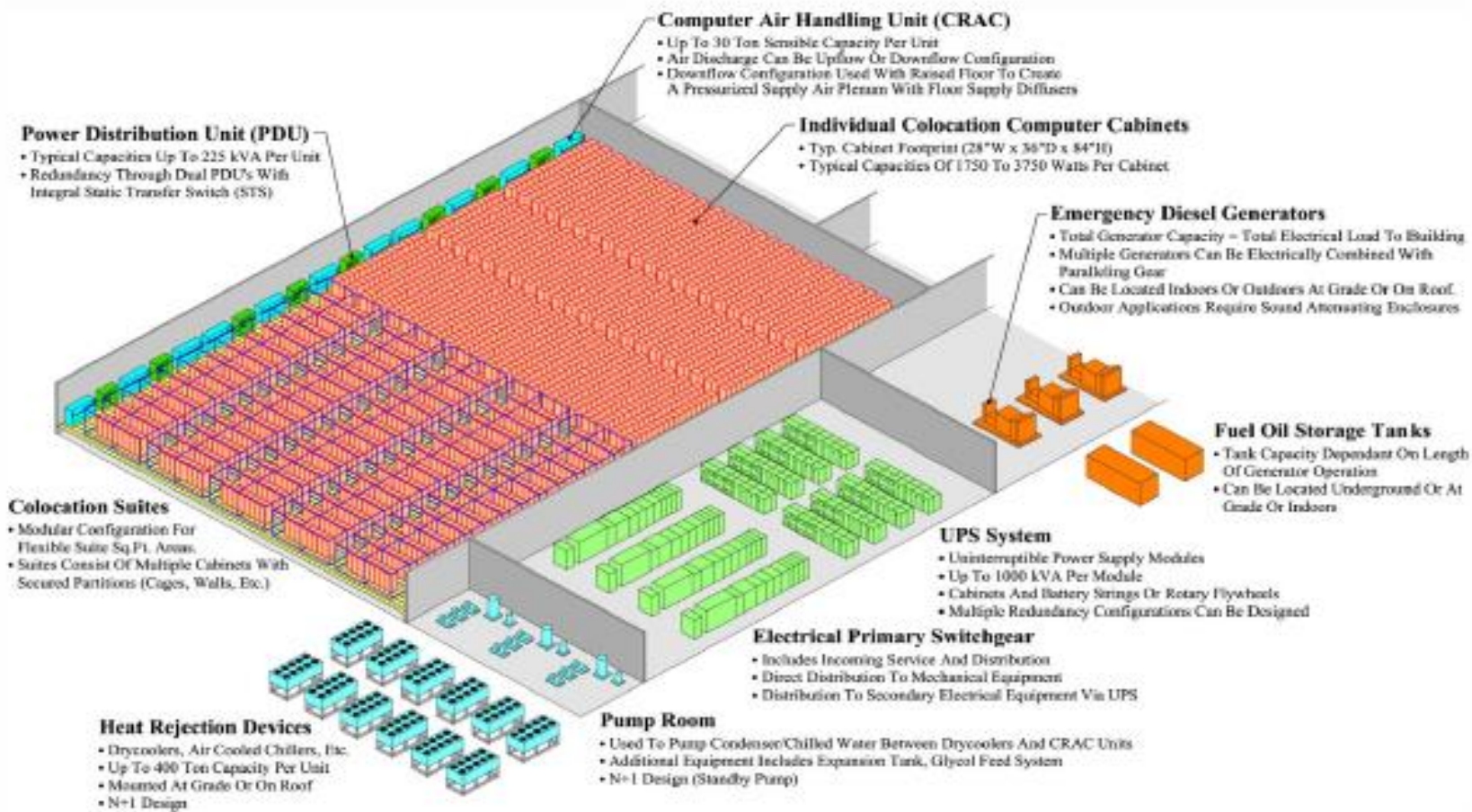


FIGURE 4.1: The main components of a typical datacenter (image courtesy of DLB Associates [23]).

- ◆ Značajan deo prostora i opereme je za napajanje i hlađenje
- ◆ Rač. kabineti za osnovni sistem + dodatne fleksibilne konfiguracije

Zahtevi napajanja i hlađenja

- ◆ Sistem hlađenja na vodu (isparavanje i izlivanje)
 - Npr. 70,000 do 200,000 galona dnevno za 8 MW postrojenje
- ◆ Rasčlanjivanje troškova napajanja:
 - Vodeni hladnjaci: 30-50% energije korišćene za IT
 - Vazdušno hlađenje: 10-20% IT energije
- ◆ Koliko servera može da podrži WSC?
 - Spec. snage na „pločici“ servera daje max potrošnju
 - Meri se snaga pod stvarnim opterećenjima
 - Predimenzionisati ukupnu snagu servera za 40%, uz striktan nadzor snage

Pojam efikasnosti Centra za pod.

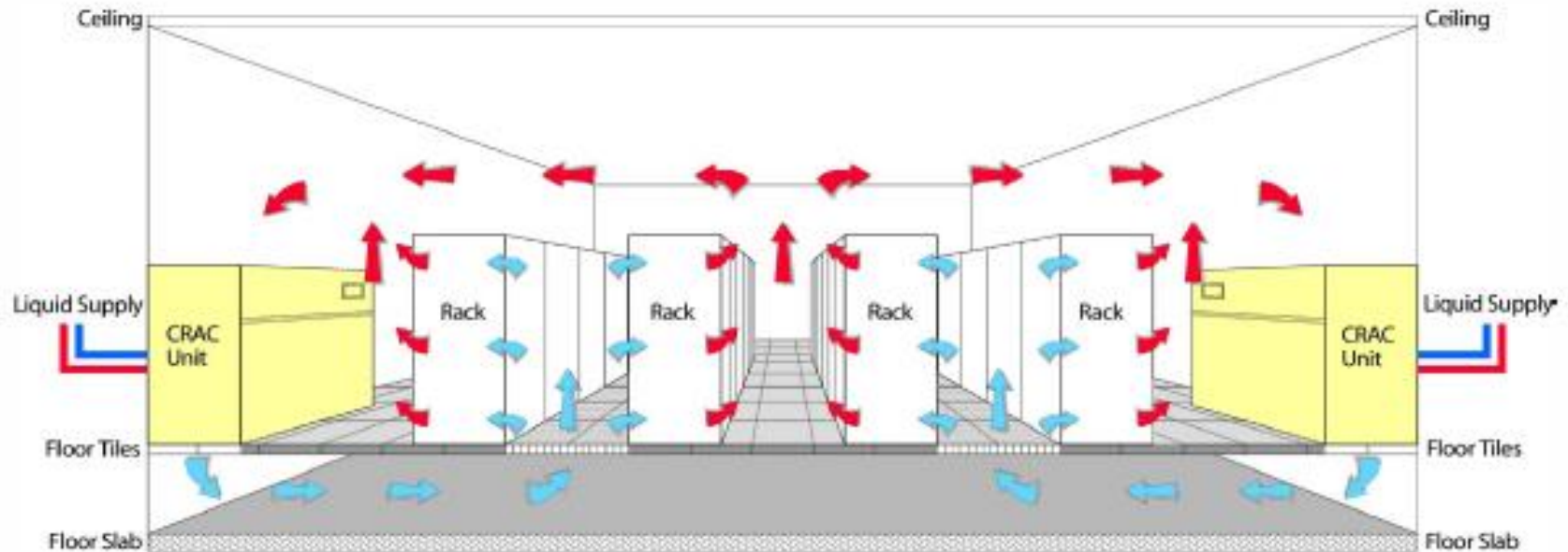


FIGURE 4.2: Datacenter raised floor with hot-cold aisle setup (image courtesy of DLB Associates [23]).

- ◆ Na slici je prikazan dovod hladnog i odvod toplog vazduha
- ◆ Efikasnost centra je (energija \Leftrightarrow snaga):

$$\text{Efikasnost} = \text{Rad} / \text{Ukupna energija} = \\ (1/\text{PUE}) \times (\text{Rad} / \text{Energija IT opreme})$$

Merenje efikasnosti WSC

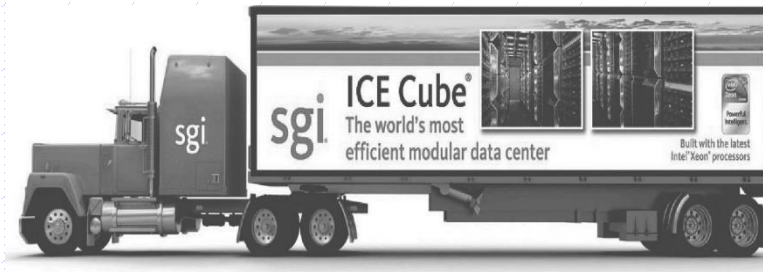
◆ Efikasnost korišćenja snage (PUE)

- = Ukupna snaga postrojenja / Snaga IT opreme
- Median PUE u studiji iz 2006 je bio 1.69

◆ Performansa

- Latentnost je važna metrika pošto je vide korisnici
- Studija Bing: korisnici će sve manje koristiti pretragu kako se vreme odziva povećava
- SLOs (Service Level Objectives) / SLAs (Service Level Agreements)
 - ◆ Npr. 99% zahteva da bude ispod 100 ms

Modularni centri za podatke



◆ Pojavili su se i mobilni centri za podatke

Računanje u oblaku

- ◆ WSC nudi skalu ekonomija koja se ne može postići sa centrom za podatke:
 - 5.7 puta manje troškova skladišta
 - 7.1 puta manje administrativne troškove
 - 7.3 puta manje troškove umrežavanja
- ◆ Ovo je omogućilo pojavu usluga oblaka kao što je AWS (Amazon Web Services)
 - “Uslužno računanje” (Utility computing)
 - Bazira se na korišćenju otvorenih VM i OS

Osnovne tehnologije za oblak

Tehnologija	Zahtevi i korist
Brzo raspoređivanje platforme	Brzo, efikasano i fleksibilno raspoređivanje resursa oblaka radi obezbeđivanja dinamičkog okruženja za računanje
Virtuelni klasteri na zahtev	Virtuelizovani klaster VM obezbeđen da zadovolji zahteve korisnika, sa mogućnošću rekonfiguracije sa promenom opterećenja
Tehnike za više stanara	SaaS za distribuiranje SW na veliki broj korisnika za njihovo istovremeno korišćenje, kao i deljenje resursa ako je to potrebno
Masivna obrada podataka	Pretraga interneta i web usluge koje obično zahtevaju masivnu obradu podataka, posebno za podršku personalizovanim uslugama
Komunikacija web skale	Podrška za aplikacije e-trgovine, obrazovanja na daljinu, telemedicine, socijalne mreže, digitalne vlade, i digitalnu zabavu
Distribuirano skladište	Skladište velike skale za personalne zapise i javne arhive, koje zahtevaju distribuirano skladište nad oblacima
Usluge licenciranja i naplate	Rukovanje licencama i usluge naplate, koje znatno koriste svim tipovima usluga oblaka u uslužnom (utility) računanju

Računanje u oblaku kao usluga

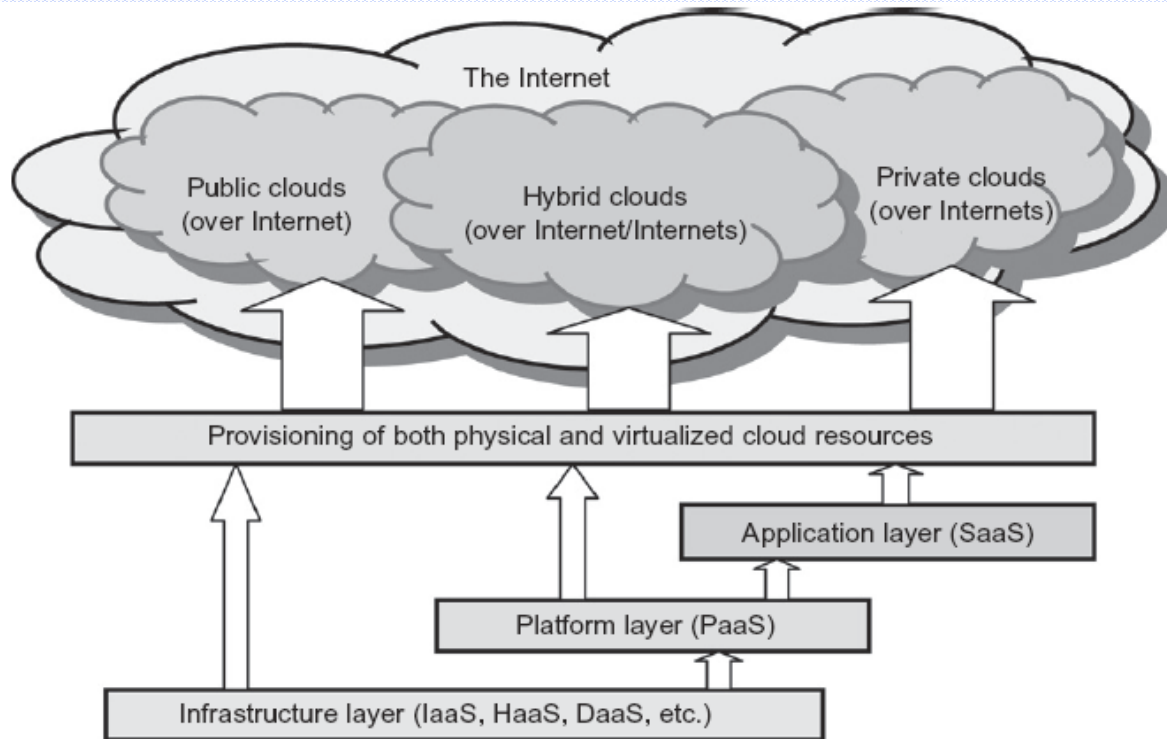


FIGURE 4.15

Layered architectural development of the cloud platform for IaaS, PaaS, and SaaS applications over the Internet.

- ◆ Servisi oblaka obezbeđuju fizičke i virtuelne resurse
- ◆ Nivoi servisa: SaaS, PaaS, IaaS

Tržišno-orijentisana arhitektura oblaka sa servisima na zahtev

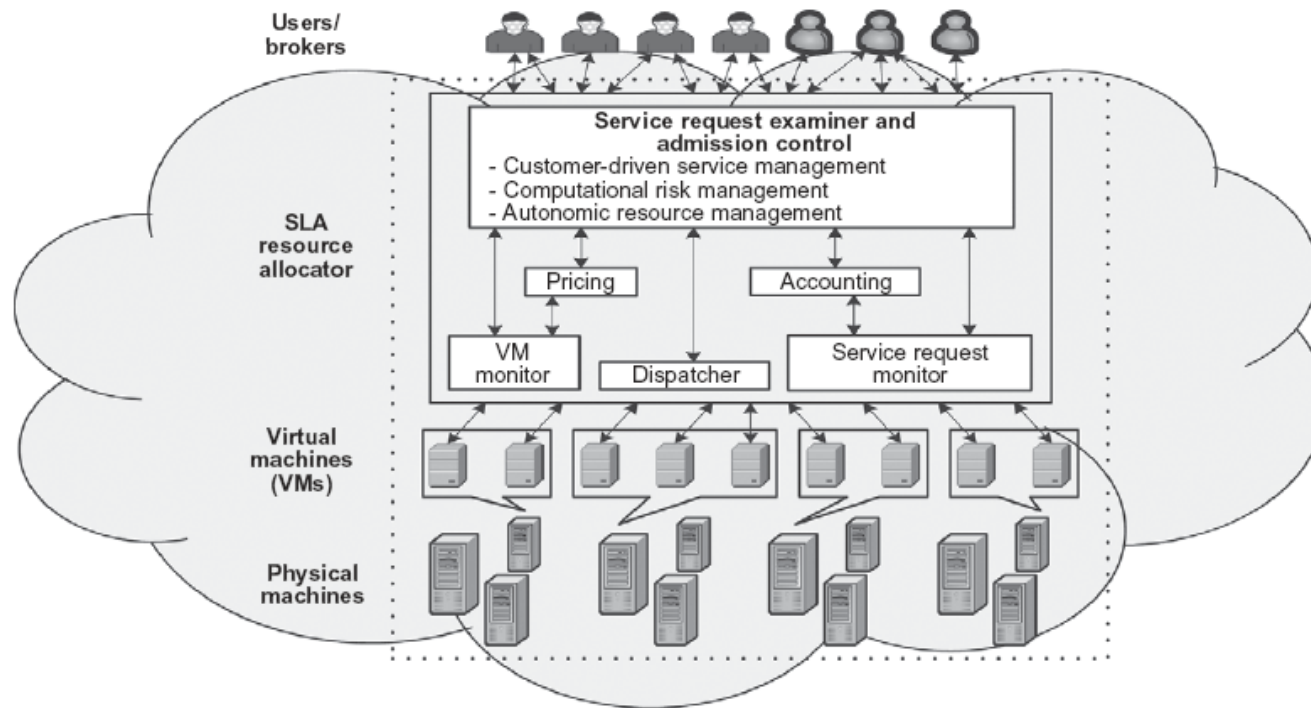


FIGURE 4.16

Market-oriented cloud architecture to expand/shrink leasing of resources with variation in QoS/demand from users.

(Courtesy of Raj Buyya, et al. [11])

- ◆ Rukovalac SLA resursima: utvrđivanje cena, obračun, monitor zahteva za servis, VM monitor i raspoređivač

Funkcije rukovaoca infrastrukturom i virtuelizovanom infrastrukt. (1/2)

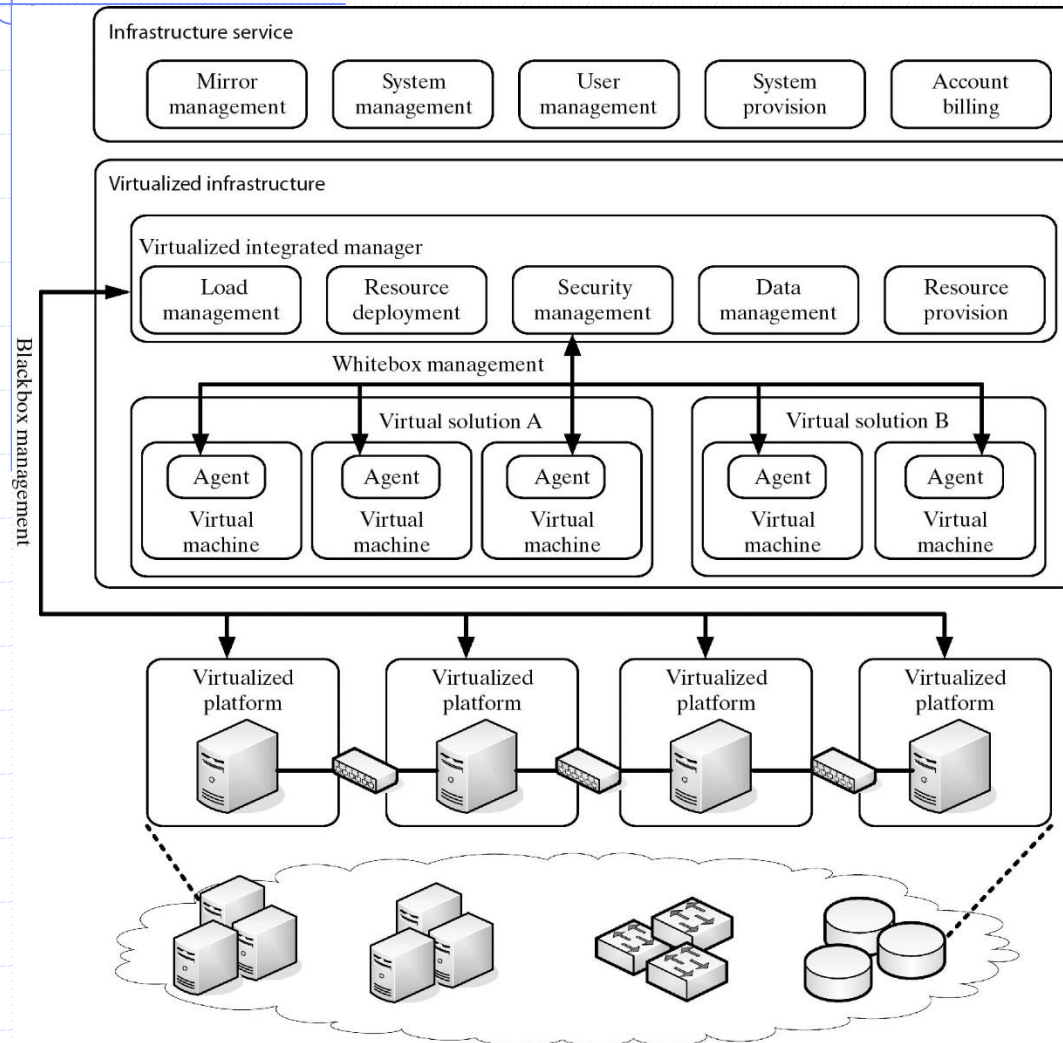
◆ Funkcije rukovaoca infrastrukturom:

- Rukovanje sistemom, korisnicima i ogledalom
- Obezbeđivanje sistema
- Obračun naplate

◆ Funkcije rukovaoca virt. infrastrukturom :

- Rukovanje opterećenjem
- Raspoređivanje resursa
- Rukovanje zaštitom
- Rukovanje podacima
- Obezbeđivanje resursa

Funkcije rukovaoca infrastrukturom i virtuelizovanom infrastruk. (2/2)



Virtuelizovani resursi u oblacima

Dobavljač	AWS	Microsoft Azure	GAE
Računski oblak sa virtuelnim klasterom servera	X86 instrukcije, Xen VM, elastičnost resursa omogućava skalabilnost kroz virt kalster, ili 3-ća strana kao RightScale obezbeđuje klaster	Izvršna VM za zajednički jezik, snabdevena kroz deklarativne opise	Zaglavlja aplikativnog okruženja pisana u Python, automatsko skaliranje na gore i dole, preuzimanje servera nezavisno od web aplikacija
Skladišni oblak sa virtuelnim skladištem	Modeli za blok skladište (EBS) i key/blob skladište (SimpleDB), skaliranje od EBS do potpuno automatskog (SimpleDB, S3)	Usluge SQL Data (ograničen pogled SQL servera), usluga Azure storage	MegaStore/BigTable
Usluge mrežnog oblaka	Deklarativna topologija na IP nivou, postavka skrivena, zaštita grupe ograničava komunikaciju, zona raspoloživosti izoluje mrežne otkaze, elastičan IP	Automatski sa korisničkim deklarativnim opisima ili ulogama aplikativnih komponenti	Fiksna topologija za web aplik. sa 3 točka, skaliranje na gore i dole je automatsko i nevidljivo za programera

Izazovi računanja u oblaku (1/2)

◆ Problemi iz industrije (Dobavljači):

- Troškovi zamene delova: Eksponencijalno povećanje troškova održavanje infrastrukture
- Zaključavanja od strane proizvođača: Nepostojanje standardong API može biti problem
- Standardizacija: Nepostojanje standardnih metrika za QoS
- Zaštita i poverljivost: Model poverenja za računanje u oblaku
- Kontrolni mehanizam: Korisnici nemaju nikakvu kontrolu nad infrastrukturama

Izazovi računanja u oblaku (2/2)

◆ Problemi iz zajednice istraživača:

- Konflikt sa nasleđenim programima: Teškoće u razvoju novih aplikacija zbog nedostatka kontrole
- Poreklo: Kako ponoviti rez. na različitim infrastrukt.
- Smanjenje latentnosti:
 - ◆ Ne koriste se posebno projektovane međuveze
 - ◆ Vrlo niska kontrolabilnost strukture međuveze zbog apstrakcije
- Model programiranja: Otežano debugiranje, Detalji infrastrukture su skriveni
- Merenje QoS: Pogotovo za sveprisutno računanje gde se kontekst menja

Javni oblaci i ponuđene usluge

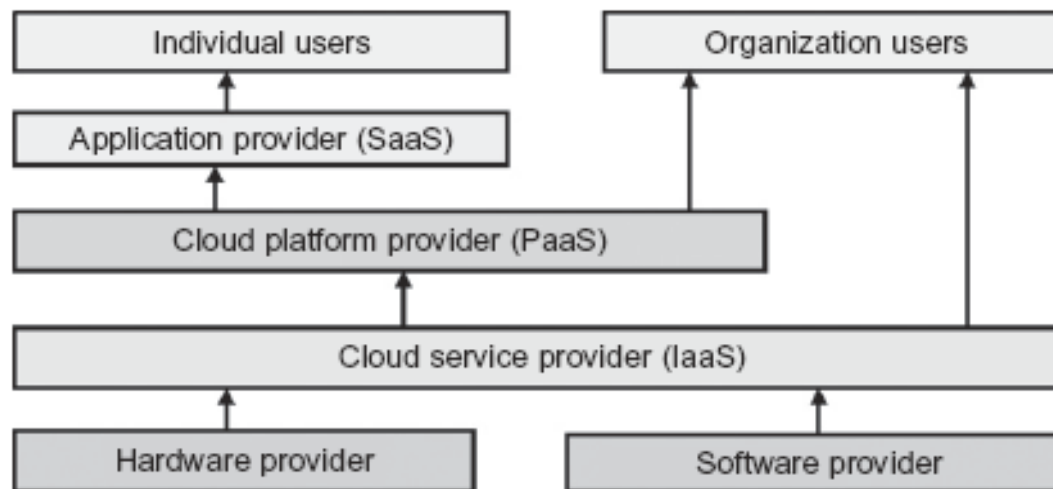


FIGURE 4.19

Roles of individual and organizational users and their Interaction with cloud providers under various cloud service models.

- ◆ Korisnici oblaka mogu biti pojedinci ili organizacije
- ◆ Pojedinci su neposredni korisnici SaaS
- ◆ Organizacije su korisnici PaaS i/ili IaaS

Pet glavnih platformi oblaka (1/2)

◆ IBM

- SaaS: Lotus live
- PaaS: BlueCloud, WCA, RC2
- IaaS: Ensembles

◆ Amazon

- IaaS: AWS

◆ Google

- SaaS: Gmail, Docs
- PaaS: App Engine (GAE)

Pet glavnih platformi oblaka (2/2)

◆ Microsoft

- SaaS: .NET servis, Dynamic CRM
- PaaS: Windows Azure
- IaaS: Windows Azure

◆ Salesforce

- SaaS: Online CRM, Gifttag
- PaaS: Force.com

PaaS: Google App Engine (1/4)

- ◆ Razvoj web app u Google centru za podatke
 - Sa automatskim skaliranjem u skladu sa zahtevima
- ◆ To je besplatna usluga do određene granice
 - Zahteva samo Gmail nalog za pristup
 - Dodatno skladište, propusni opseg, i sati rada instance se naplaćuju
- ◆ Podržava Java, Python i Go programske jezike
- ◆ Kvalitet posluživanja:
 - Sve App koje se plaćaju imaju 99.95% nivo SLA
 - App Engine je projektovan da izdrži višestruke ispade centara za podatke bez zastoja

PaaS: Google App Engine (2/4)

◆ GAP ograničenja:

- Može da izvršava samo kod pozvan iz HTTP zahteva
- Java aplikacije mogu koristiti samo podskup iz JRE standardne edicije
- Java aplikacije ne mogu stvarati nove niti

◆ Informacije o GAE:

- Prednja str.: <http://code.google.com/appengine/>
- SDK: <http://code.google.com/appengine/download>

PaaS: Google App Engine (3/4)

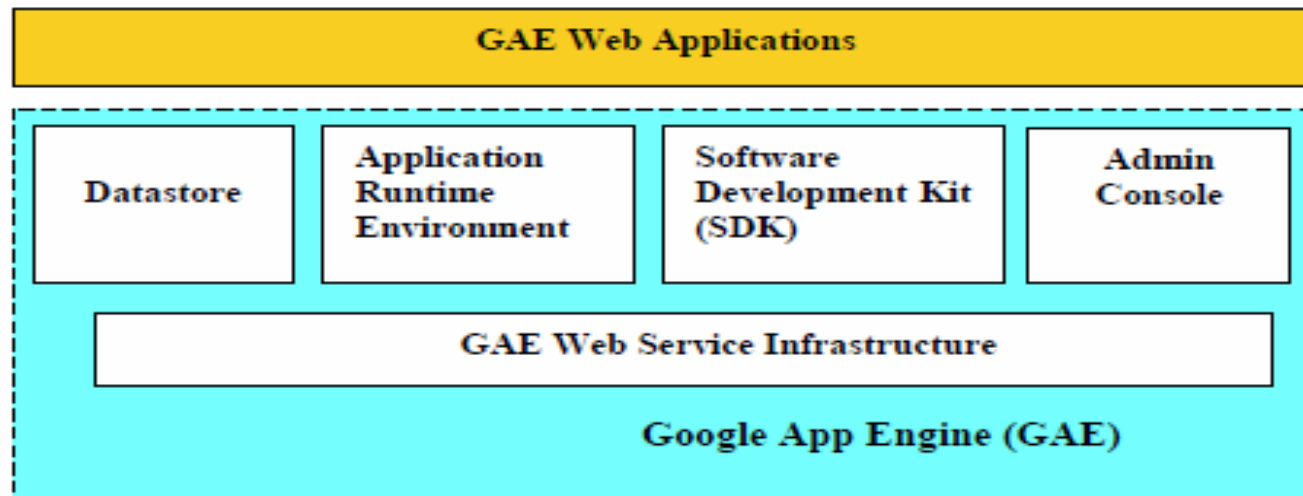


Figure 7.24 Functional components in the Google App Engine (GAE)
(Courtesy of Google, <http://code.google.com/appengine/>)

- ◆ Funkcionalne komponente GAE u tri sloja:
 - GAE aplikacije
 - Skladište podataka, izvršno okruženje, SDK, admin konzola
 - Infrastruktura GAE servisa

PaaS: Google App Engine (4/4)



FIGURE 4.20

Google cloud platform and major building blocks, the blocks shown are large clusters of low-cost Servers.

(Courtesy of Kang Chen, Tsinghua University, China)

◆ Arhitektura Google oblaka:

- Spec čvorovi: Raspoređivač, Server za brave (Chubby), GFS master
- Običan čvor: MapReduce posao, BigTable server, prateći raspoređivač, Server GFS blokova, OS Linux

IaaS: AWS (1/3)

◆ Servisi:

- EC2 (Elastic Compute Cloud): omogućava korisnicima da iznajme virtualne računare radi puštanja svojih aplikacija. Omogućava skalabilan raspored.
- S3 (Simple Storage Service): pruža uslugu objektno-orijentisanog skladišta.
- EBS (Elastic Block Service) obezbeđuje blok skladište za podršku tradicionalnih aplikacija.
- Amazon DevPay je jednostavna usluga za naplate i rukovanje računima, radi lakšeg vođenja biznisa.

IaaS: AWS (2/3)

◆ Servisi - nastavak:

- MPI klaster: koristi HW virtuelizaciju, i korisnici mogu da prave nove AMI (Amazon Machine Image)
- AWS import/export: omogućava isporuku velike količine podataka u i iz EC2 isporukom fizičkih diskova
- Sistemi posredovanja (Brokering): model za kontrolu senzora i podršku za pametne telefone i tablete
- Male kompanije mogu da postave poslovanje na AWS.

IaaS: AWS (3/3)

Compute Amazon Elastic Compute Cloud (EC2) Amazon Elastic MapReduce Auto Scaling	Messaging Amazon Simple Queue Service (SQS) Amazon Simple Notification Service (SNS)	Storage Amazon Simple Storage Service (S3) Amazon Elastic Block Storage (EBS) AWS Import/Export
Content Delivery Amazon CloudFront	Monitoring Amazon CloudWatch	Support AWS Premium Support
Database Amazon SimpleDB Amazon Relational Database Service (RDS)	Networking Amazon Virtual Private Cloud (VPC) Elastic Load Balancing	Web Traffic Alexa Web Information Service Alexa Top Sites
E-Commerce Amazon Fulfillment Web Service (FWS)	Payments & Billing Amazon Flexible Payments Service (FPS) Amazon DevPay	Workforce Amazon Mechanical Turk

- ◆ Grupe servisa: računanje, isporuka sadržaja, bazapodataka, e-trgovina, poruke, nadzor, umrežavanje, plaćanje i obračun, skladište, podrška, web saobraćaj, radna snaga

Amazon lekcija

- ◆ Zastoj u trajanju od 3 dana od 22.04.2011.
- ◆ 1000x biznisa van rada. Npr. Pfizer, Netflix, Quora, Foursquare, Reddit
- ◆ SLA ugovor je bio na:
 - 99.95% raspoloživost (<4.5h zastoja)
 - 10% penali, inače

CNNMoney
A Service of CNN, Fortune & Money

FORTUNE

HomeVideoBusiness NewsMarketsTerm SheetEcon

Why Amazon's cloud Titanic went down



By David Goldman, staff writer April 22, 2011: 5:37 PM ET

NEW YORK (CNNMoney) -- This was never supposed to happen.

Amazon Web Services is the Titanic of cloud hosting, designed with backups to the backups' backups that prevent hosted websites and applications from failing.

PaaS: Microsoft Azure (1/3)

◆ Model:

- Azure koristi centar za pod. koji rukuje resursima za računanje i skladištenje
- Azure Development Kit (ADK) se može koristiti za izvršenje lokalne verzije Azure na Windows 7
- Sve usluge oblaka mogu da interaguju sa MS web app: Windows Live, Office Live, Exchange Online, itd.

PaaS: Microsoft Azure (2/3)

◆ Model - nastavak:

- Nudi platformu oblaka zasnovanu na Windows
- Aplikacije se prave na VM koja se oslanja na usluge centra za podatke
- Azure rukuje svim serverima, skladištem i mrežnim resursima centra za podatke

PaaS: Microsoft Azure (3/3)

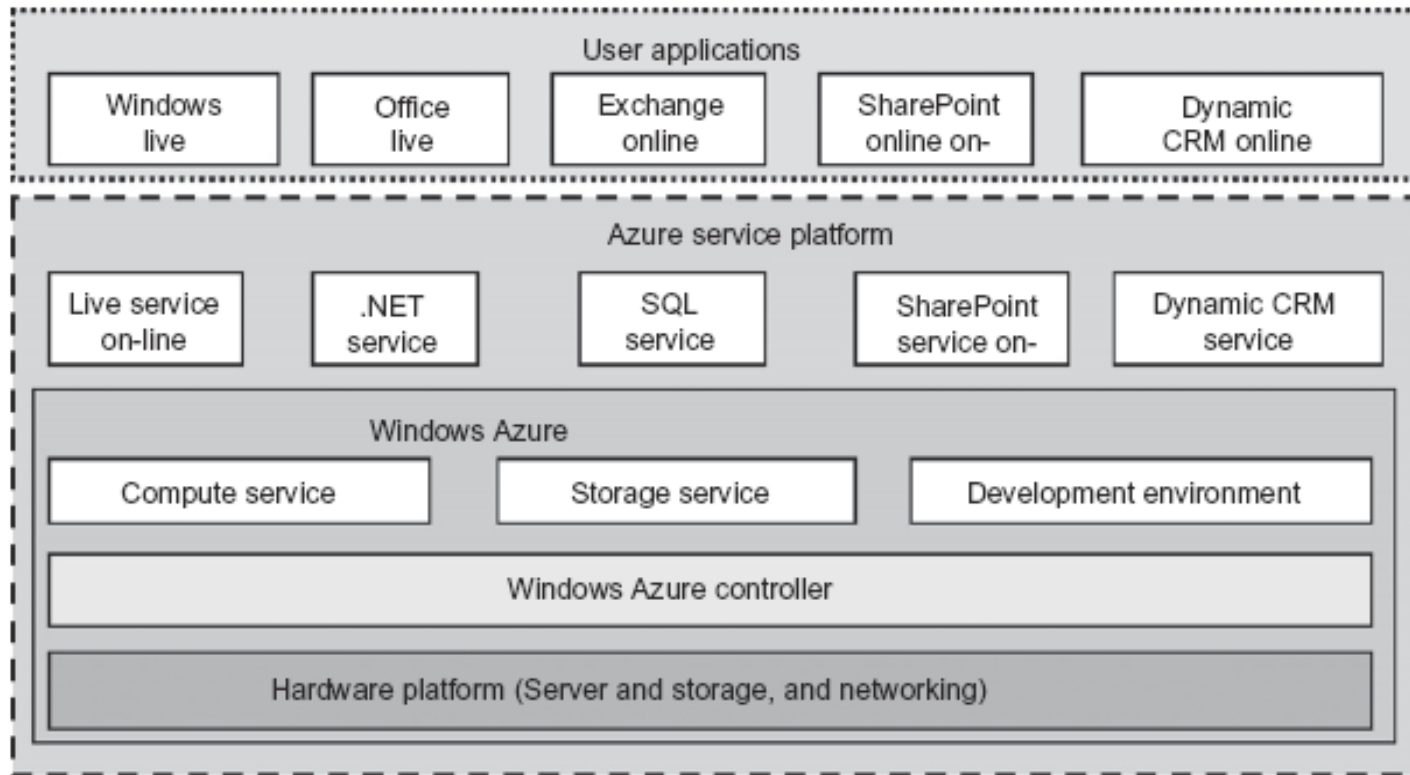


FIGURE 4.22

Microsoft Windows Azure platform for cloud computing.

(Courtesy of Microsoft, 2010, <http://www.microsoft.com/windowsazure>)

- ◆ Eksterni servisi: Live, .NET, SQL, SharePoint, Dynamic CRM
- ◆ Unutrašnji servisi: računanje, skladište, razvojno okruženje
- ◆ Azure kontroler rukuje hardverom

Razlike između oblaka iz ugla dobavljača, prodavca i korisnika

Razlike između oblaka iz perspektive dobavljača, prodavca i korisnika			
Uloga	IaaS	PaaS	SaaS
IT administratori / dobavljači oblaka	Nadzor SLA	Nadzor SLA i omogućavanje usluga platforme	Nadzor SLA i razmeštanje SW
Programeri (prodavci)	Razmeštanje i skladištenje podataka	Omogućavanje platformi kroz konfiguratore i API-je	Razvoj i razmeštanje SW
Krajnji korisnici ili Biznis korisnici	Razmeštanje i skladištenje podataka	Razmeštanje i testiranje web SW	Korišćenje biznis SW

Skladišne usluge u sistema za računanje u oblaku

Skladišne usluge u sistema za računanje u oblaku

Skladišni sistem	Osobine
GFS (Google File System)	Vrlo velik i izdrživ propusni opseg za čitanje i pisanje, uglavnom kontinualan pristup umesto slučajnog pristupa. Programska sprega slična POSIX sprezi za pristup sistemu datoteka.
HDFS (Hadoop Distributed File System)	Otvoreni klon GFS-a. Napisan u Java. Programska sprega slična POSIX ali nije identična.
Amazon S3 i EBS	S3 služi za dobavljanje/skladištenje podataka sa/na udaljene servere. EBS je izgrađen iznad S3 za korišćenje virtuelnih diskova radi izvršavanja EC2 instanci.

Zaštita i Barijere poverenja za računanje u oblaku

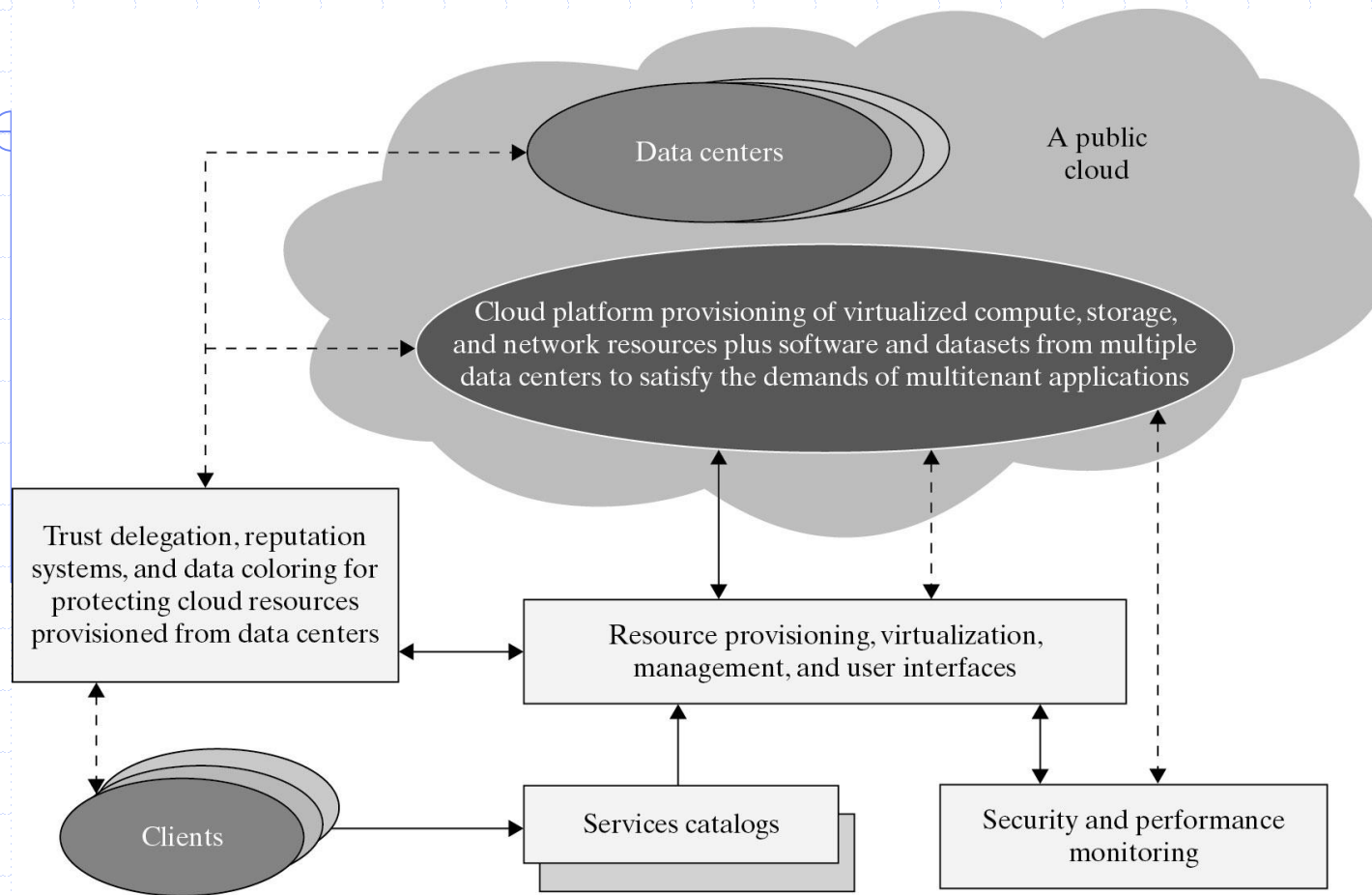
◆ Ciljevi sistem zaštite centara za podatke:

- Zaštita resursa oblaka i podrška privatnost korisnika i integritetu podataka

◆ Tehnike:

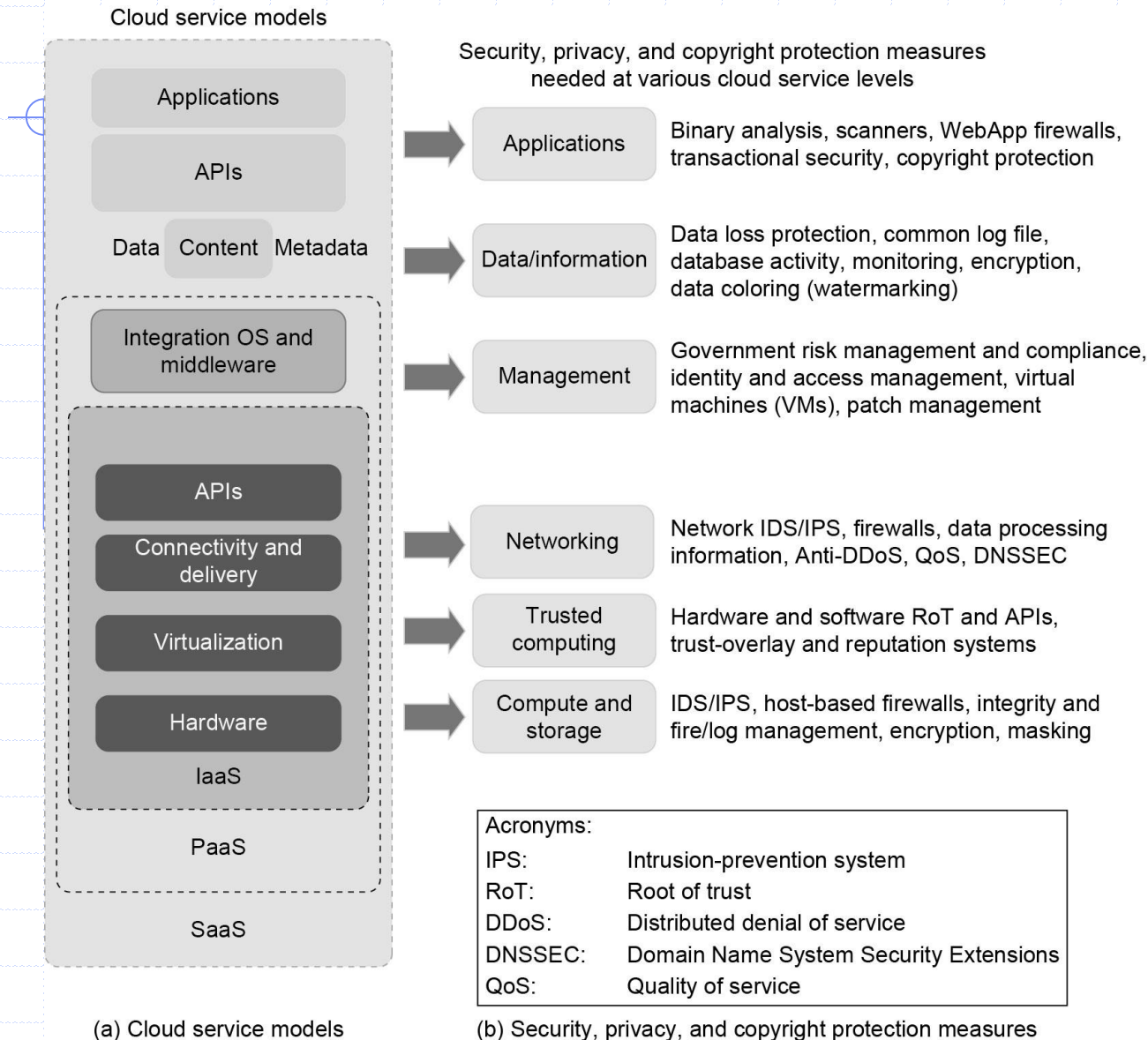
- Prekrivačka mreža: za sistem sa reputacijom radi uspostavljanja poverenja među centrima za podatke
- Tehnika vodenog žiga se predlaže za zaštitu deljenih objekata podataka i masivno distriburanih SW modula
- Ove tehnike zaštićuju autentifikaciju korisnika i zatežu kontrolu pristupa podacima u javnim oblacima
- Novi pristup može biti isplativiji od tradicionalnog šifrovanja i zaštitnih barijera radi zaštite oblaka

Platforma oblaka svesna zaštite



- ◆ Klijenti koriste delegiranje poverenja, sistem reputacije, i bojenje podataka
- ◆ Sistem obezbeđuje nadzor zaštite i performanse

Modeli servisa oblaka i mere zaštite



◆ SaaS: zaštita aplikacija i zaštita podataka/info.

◆ PaaS: rukovanje rizikom i usaglašenošću, identitetom i pristupom, VM, zakrpama

◆ IaaS: umrežavanje, poverljivo računanje, računanje i skladište