

Optimizacija troškova prilikom korišćenja AWS platforme

Ivo Martinović, Prof. Dr Mirjana Radivojević

Sadržaj - U ovom radu predstavljen je *Amazon Web Services* platforma, kao i servisi koje ona pruža. Oni su grupisani, u zavisnosti od vrste usluge koju pružaju, u četiri glavne grupe: Računarski servisi, servisi za baze podataka, servisi za skladištenje i mrežni servisi. Jedan od najvećih izazova prilikom korišćenja ove platforme je optimizacija troškova koje servisi stvaraju. Rad detaljno prolazi kroz troškove servisa, protokole kako ih smanjiti i kako to funkcioniše na realnom primeru u kojem se detaljno porede se optimizovana i neoptimizovana infrastruktura.

Ključne reči - AWS, Amazon Web Services, Cloud, Infrastruktura, Troškovi, Optimizacija

I. UVOD

Infrastruktura¹ u informacionoj tehnologiji predstavlja temelj i skup ključnih resursa neophodnih za funkcionisanje sistema i pružanje različitih tehnoloških usluga. Njena uloga je ključna u podršci poslovnih procesa i aktivnosti organizacije, pružajući stabilnost, sigurnost, i efikasnost u korišćenju različitih resursa. Cilj infrastrukture je omogućiti pouzdan i održiv rad svih tehnoloških aspekata organizacije.

Lokalna infrastruktura, tj. tradicionalno održavanje i upravljanje fizičkim računarima, serverima i mrežnom opremom unutar organizacije, može se suočiti sa nizom izazova² kao što su:

- Visoki troškovi investicije: Izgradnja i održavanje lokalne infrastrukture zahteva značajne početne investicije u hardver, softver, mrežnu opremu i osoblje,
- Ograničena skalabilnost: Lokalna infrastruktura može biti ograničena u smislu skalabilnosti, posebno ako organizacija brzo raste ili ima promenljive potrebe za resursima,
- Složeno upravljanje i održavanje: Održavanje lokalne infrastrukture može biti složeno i zahteva stručno osoblje za rešavanje problema, nadogradnje sistema i implementacije sigurnosnih mehanizama,
- Rizik od hardverskih kvarova: Hardverski kvarovi mogu uzrokovati značajne zastoje u radu, naročito ako organizacija nema efikasne mere zaštite podataka i oporavak od katastrofe.
- Nedovoljna bezbednost podataka: Osiguranje podataka u lokalnom okruženju može biti izazovno, posebno u pogledu praćenja i primene najnovijih sigurnosnih standarda.

II. RAČUNARSTVO U OBLAKU

*Cloud computing*³ predstavlja model pružanja računarskih resursa putem interneta krajnjim korisnicima. Umesto da organizacije održavaju sopstvenu infrastrukturu, softver i resurse, oni ih iznajmljuju ili koriste usluge koje se nalaze na udaljenim serverima. Prednosti ovog modela su:

- Smanjenje troškova: Ne postoji potreba za kupovinom i održavanjem hardverske opreme i tako se smanjuju kapitalni troškovi,
- Skalabilnost: Mogućnost brzog povećanja ili smanjenja resursa prema potrebama,
- Visoka dostupnost: *Cloud* provajderi često obezbeđuju visok nivo dostupnosti i redundantnosti.

Postoje različite vrste računarstva u oblaku³ koje pružaju različite nivoe resursa, upravljanja i privatnosti, a to su:

- Javni Oblak: Provajder omogućava pristup resursima kao što su aplikacije, skladišta za podatke i drugi resursi dostupni za javnost, nezavisno da li se radi o pojedincima ili organizacijama, putem interneta. Usluge mogu biti besplatne ili se koristi model plaćanja po korišćenju. Infrastruktura se nalazi u vlasništvu provajdera i nije dostupna za uvid ili kontrolu korisnicima. Ova vrsta *cloud-a* pruža visok stepen skalabilnosti, ali korisnici dele resurse sa drugim organizacijama,
- Privatni Oblak: Namenjen je isključivo jednoj organizaciji. Infrastruktura može biti smeštena unutar organizacije ili kod provajdera, ali je namenjena samo internim korisnicima organizacije. Privatni *cloud* omogućava veći stepen kontrole i prilagodljivosti, ali može biti skuplji za održavanje,
- Hibridni Oblak: Predstavlja kombinaciju privatnog i javnog i tako omogućava zajedničko korišćenje istih. Koristi se ukoliko je potrebno proširiti infrastrukturu organizacije na neko određeno vreme,
- Zajednički Oblak: Predstavlja vrstu *cloud-a* koji deli nekoliko organizacija. Infrastruktura podržava posebne zajednice koje imaju zajedničke potrebe, misije, zahteve sigurnosti i drugo. Njima mogu upravljati same organizacije ili neko drugi (provajder usluga). Ovde se troškovi dele između manjeg broja klijenata tako da su mogućnosti uštede ograničene.

III. AMAZON WEB SERVICES

*Amazon Web Services*⁴ (AWS) je skup *cloud computing* usluga koje pruža Amazon. Prve usluge za korišćenje su u javnost pustili 2006. godine. Danas broje preko dve stotine servisa, nekoliko miliona korisnika i na desetine hiljada partnera. Kao i ostali provajderi, AWS koristi *pay-as-you-go* model gde korisnici plaćaju ono što koriste. Servise možemo podeliti u nekoliko grupa:

- Računarski servisi⁵ (*Computing Services*): Ove usluge nude skalabilne i fleksibilne resurse, omogućujući korisnicima da prilagode svoje računarske potrebe prema trenutnim zahtevima njihovih aplikacija. Tu spadaju:
 - Amazon EC2 (*Elastic Compute Cloud*): Pruža skalabilne virtuelne servere. Korisnici mogu odabrati različite vrste instanci s različitim resursima, poput procesorske snage, memorije i mrežnih performansi,
 - AWS Lambda: Predstavlja uslugu za *serverless computing*, gde korisnici mogu izvršavati funkcije (delove koda) bez potrebe za upravljanjem infrastrukturom. Lambda automatski skalira resurse prema potrebi,
 - Amazon ECS (*Elastic Container Service*): Omogućava korisnicima upravljanje kontejnerima. To je korisno za implementaciju mikroservisa i upravljanje njihovim životnim ciklusom,
 - Amazon EKS (*Elastic Kubernetes Service*): Pruža potpunu kontrolu nad Kubernetes klasterima, što olakšava implementaciju, upravljanje i skaliranje kontejnerskih aplikacija,

- Amazon Lightsail: Jednostavnija usluga koja olakšava pokretanje i upravljanje virtuelnim privatnim serverima, web aplikacijama i bazama podataka. Pogodna je za lične projekte i studente koji žele brzo postaviti i pokrenuti svoje aplikacije.
- Servisi za skladištenje (*Storage Services*): Ovi servisi omogućavaju korisnicima da skladište, upravljaju i pristupaju podacima koji se nalaze na infrastrukturi provajdera. Ovde pripadaju:
 - Amazon S3 (*Simple Storage Service*): Skalabilna i izdržljiva usluga za skladištenje podataka. Amazon S3 omogućava korisnicima da skladište i preuzimaju podatke, a takođe podržava različite funkcionalnosti kao što su verzionisanje, enkripcija i upravljanje pravima pristupa,
 - Amazon EBS (*Elastic Block Store*): Ovo je servis za skladištenje blokova podataka koji se koristi za povezivanje sa virtuelnim mašinama. Amazon EBS pruža perzistentno skladište koje može biti vezano za bilo koju EC2 instancu u istom regionu,
 - Amazon Glacier: Ova usluga je dizajnirana za dugoročno i jeftino skladištenje podataka koji se retko koriste. Amazon Glacier omogućava arhiviranje podataka po niskim cenama, ali pruža i sporiji pristup podacima u poređenju sa Amazon S3,
 - AWS Storage Gateway: Servis koji omogućava povezivanje lokalnih mašina sa fajl sistemima koje održava provajder. Ova usluga olakšava integraciju lokalnih sistemskih okruženja sa provajderom i omogućava uporedno korišćenje lokalnih i udaljenih resursa. Cilj ovog servisa je da omogući jednostavan i efikasan način za prenos podataka između lokalnih mašina i udaljenih skladišta,
 - AWS Backup: Usluga koja omogućava korisnicima da jednostavno i efikasno upravljaju rezervnim kopijama svojih podataka. Cilj ovog servisa je da olakša kreiranje, upravljanje i praćenje rezervnih kopija za različite resurse, uključujući Amazon EBS (*Elastic Block Store*), Amazon RDS (*Relational Database Service*) baze podataka, Amazon DynamoDB tabele, Amazon EFS (*Elastic File System*) sisteme i druge.
- Mrežni servisi⁵ (*Networking Services*) : Podrazumevaju AWS proizvode koji omogućavaju konfiguraciju, upravljanje i optimizaciju mreže koju koriste ostali servisi korisnika. Neki od njih su:
 - Amazon VPC (*Virtual Private Cloud*): Amazon VPC omogućava korisnicima da virtuelno izoluju svoje resurse kao i da kreiraju privatne mreže na infrastrukturi provajdera. Korisnici mogu definisati svoje IP adrese, podmreže, rute i konfigurirati ulaze u mrežu,
 - Amazon Route 53: Ovaj servis pruža skalabilno i pouzdano upravljanje DNS (*Domain Name System*). Amazon Route 53 omogućava registraciju i upravljanje domenima, kao i rukovanje DNS upitima kako bi se postigla visoka dostupnost i otpornost na greške,
 - AWS Direct Connect: Ovaj servis omogućava korisnicima direktnu vezu sa svojim AWS resursima, izbegavajući internet. To omogućava privatnu, sigurnu i brzu konekciju između korisnika i provajdera,
 - Elastic Load Balancing (ELB): Pruža automatizovanu distribuciju saobraćaja na više različitih krajnjih tačaka kako bi se postigla bolja dostupnost i otpornost. Takođe, pruža i različite vrste balansiranja opterećenja, uključujući *Application Load Balancer* i *Network Load Balancer*,

- CloudFront: Predstavlja Amazonovo rešenje za brže i efikasnije dostavljanje multimedijalnog sadržaja krajnjim korisnicima (*Content Delivery Network*). Svaki put kada se podaci zatraže iz Amazonove mreže, oni se keširaju na obodima mreže i onda im se kasnije brže pristupa. Ovo omogućava bolje performanse aplikacija i softvera nezavisno od toga u kom regionu se oni nalaze.
- Servisi za baze podataka⁵ (*Database servisi*): Ovi servisi omogućavaju korisnicima da brzo i lako podignu, upravljaju, skaliraju i održavaju baze podataka na infrastrukturi provajdera. Tu pripadaju:
 - Amazon RDS (*Relational Database Service*): Amazon RDS je usluga koja omogućava lako postavljanje, upravljanje i skaliranje relacionih baza podataka. Podržava popularne baze podataka poput *MySQL*, *PostgreSQL*, *MariaDB*, *Oracle* i *Microsoft SQL Server*. Amazon RDS automatski upravlja zadacima poput rezervnih kopija, ažuriranja i skaliranja resursa,
 - Amazon DynamoDB: DynamoDB je servis koji pruža korišćenje NoSQL baza podataka koje imaju visoku dostupnost, skalabilnost i brzinu. Ova baza podržava razne tipove podataka, uključujući klasične ključ-vrednost parove, dokumente i *time-series* podatke. DynamoDB se lako integriše s drugim AWS uslugama i omogućava automatsko skaliranje kapaciteta,
 - Amazon Aurora: Amazon Aurora je visoko performantna i skalabilna relacijska baza podataka kompatibilna s *MySQL-om* ili *PostgreSQL-om*. Iako podržava druge relacione baze, ovo je posebno rešenje Amazona. Pruža visoku dostupnost i otpornost na greške, a automatsko širenje kapaciteta čini je pogodnom za aplikacije s promenljivim opterećenjem. U praksi se pokazala kao bolja verzija Amazon RDS servisa, koji naravno dolazi sa većom cenom,
 - Amazon Redshift: Amazon Redshift je usluga za upravljanje analitičkim bazama podataka zasnovana na *columnar storage* arhitekturi. Idealna je za analizu velikih količina podataka i podržava SQL upite. Redshift se lako integriše s drugim AWS servisima i popularnim analitičkim alatima,
 - Amazon DocumentDB: Ovaj servis predstavlja Amazonovo rešenje za MongoDB nerelacionu bazu podataka. Prilikom kreiranja ovog servisa, bira se isključivo vrsta instance (koliko će imati RAM memorije i vCPU jezgara) dok se skladište dinamično povećava u zavisnosti od naših potreba.

IV. MODELI PLAĆANJA USLUGA

AWS pruža par načina plaćanja^{6,7} i to su:

- *Pay as you go (on demand)* model: Predstavlja plaćanje isključivo resursa koji se koriste. Ne zahteva ugovore i sve troškove obračunava u realnom vremenu, na primer: korišćenje servisa sa računarsku snagu (EC2) se naplaćuje po satu, dok na primer servisi za skladištenje (S3) trošak stvaraju na osnovu količine podataka koje sadrže.
- *Spot Instances model*: Ovaj način pruža velike popuste na resurse (čak i do devedeset posto) da bi se iskoristili nekorisćeni resursi. U slučaju da neki drugi korisnik iznajmi iste resurse za veću količinu novca, mi gubimo prava na te resurse kao i sve podatke koji su se nalazili na istim. Pre toga, dobijamo upozorenje, a onda u roku od dva minuta gubimo tu instancu. Ovaj model ne garantuje uptime i pogodan je za nekritične operacije.

- *Reserved Instances (rezervisani resursi)* model: Omogućava korisnicima da zakupe određene resurse na period od jedne ili tri godine. Jeftinija je opcija od *on demand* resursa, ali je potrebno unapred platiti servise koji će se koristiti.

V. TROŠKOVI

Troškovi prilikom korišćenja računarskih servis

- *Elastic Cloud Computing (EC2) Servis*^{6,7,8}
 - Lokacije: Možemo izabrati u kojem regionu, lokalnoj zoni ili *wavelength* zoni će se nalaziti naša EC2 mašina. Ako biramo po regionu, cena je najmanja, dok je pri biranju *wavelength* zone najveća. Takođe, lokacije gde postoje više zona će biti jeftinije od lokacija gde ima manje zona,
 - Operativni sistemi: Naša EC2 instanca može koristiti bilo koji operativni sistem. *Linux* operativni sistemi su najjeftiniji, dok su *MacOS* najskuplji,
 - Broj virtuelnih procesora: Korisnicima je dozvoljeno da biraju broj virtuelnih procesora. Jedan virtuelni procesor jednak je polovini jednog jezgra fizičkog procesora. Veći broj virtuelnih procesora stvara veći trošak,
 - Vrste instance: Različiti zahtevi korisnika zahtevaju različitu infrastrukturu. Svaka instanca stvara različit trošak. Možemo ih podeliti na: generalne instance, računarski optimizovane, memorijski optimizovane, instance optimizovane za skladištenje i računatski ubrzane instance,
 - Transfer podataka:
 - Svaki transfer koji se upisuje u instancu sa interneta je besplatan,
 - Svaki transfer u okviru Amazonove mreže, između različitih regiona se naplaćuje,
 - Svaki transfer u okviru jednog regiona se naplaćuje, ako se komunicira između različitih zona dostupnosti,
 - Svaki transfer u okviru jedne zone dostupnosti je besplatan.
 - Javna IP adresa: Svaka javna IP adresa koja je povezana na našu EC2 instancu, se naplaćuje.
- *AWS Lambda*^{6,7,8} je servis koji omogućava izvršavanje koda bez brige o infrastrukturi. Ovo je *serverless* servis koji se sam skalira po našim potrebama. Troškovi se stvaraju isključivo kada se kod izvršava i zavise od:
 - Regiona: Kao i ostali servisi, u nekim regionima će cena biti manja u odnosu na neke druge regione. To zavisi od broja data centara i servera u regionima,
 - Arhitekture: Postoje x86 i Arm arhitekture. Arm arhitektura je jeftinija od x86, ali trošak najviše zavisi od količine radne memorije,
 - Radne memorije: Možemo dodeliti bilo koji broj od 128 MB do 10240 MB. Količina radne memorije utiče na trošak koji generiše vreme u kojem se kod aplikacije izvršava,
 - Broj zahteva za izvršavanje: Svaki HTTP ili drugi zahtev da se kod izvrši se naplaćuje na million zahteva. Cena je fiksna i zavisi od regiona. Prvih milion zahteva mesečno je besplatno,
 - HTTP odgovora: Kod koji se izvršava u *Lambdi* može vraćati HTTP odgovore. Odgovori do 6 MB su besplatni, dok se ostatak plaća po fiksnoj ceni po GB. Prvih 100 GB mesečno,

koji se generišu kada HTTP odgovor pređe veličinu od 6 MB, je besplatno. Ako je HTTP odgovor 7 MB, u trošak će ući samo 1 MB (7 MB - 6 MB),

- Transfer podaka u i iz *Lambda*: Sav saobraćaj koji ulazi u *Lambda* ne stvara trošak, dok se saobraćaj koji izlazi iz *Lambda* računa po istoj ceni kao i EC2 instance. Razmena podataka između *Lambda* i drugih servisa u istom regionu je besplatna,
- *Lambda@Edge*: Predstavlja postavljanje *Lambda* koda bliže korisnicima. Cena zahteva i vreme izvršavanja stvaraju veći trošak,
- Vreme izvršavanja koda: Vreme izvršavanja se računa u jedinici mere GB/sekundi. Broj megabajta radne memorije se deli sa 1024 i to se množi sa brojem sekunda u toku kojih je naša aplikacija bila aktivna. Prvih 400000 GB/sekundi u mesecu je besplatno,
- Privremeno skladištenje: Ako je našoj aplikaciji potrebno skladište prilikom izvršavanja, ono se takođe naplaćuje. 512 MB je besplatno. Cena skladišta se računa na isti način kao i vreme izvršavanja i jedinica mere je GB/sekund,
- Obezbeđena konkurentnost: Možemo obezbediti konkurentno izvršavanje našeg koda. Ova opcija odvaja i priprema određeni broj okruženja sa određenom random memorijom. Pruža i do sedamdeset posto manju cenu izvršavanja, ali se ona dodatno plaća po broju konkurentnih okruženja, takođe u GB/sekundi.

■ Primer:

Naša aplikacije koristi 1536MB radne memorije, ima 3 miliona zahteva mesečno i prosečni zahtev traje 120ms.

$3 \text{ miliona zahteva} * 120 \text{ ms} = 360000 \text{ sekunde}$

$360000 \text{ sekunde} * 1536\text{MB}/1024/\text{MB} = 360000 * 1.5 = 540000 \text{ GB/sekundi}$

$540000 \text{ GB/sekundi} - 400000 \text{ besplatnih GB/sekundi} = 140000 \text{ GB/sekundi}$

$140000 * \$0.0000166667 \text{ (cena GB/sekundi)} = \2.33

Cena za zahteve:

$3 \text{ miliona} - 1 \text{ milion besplatnih zahteva} = 2 \text{ miliona zahteva}$

$2 \text{ miliona zahteva} * \$0.2 \text{ (cena za milion zahteva)} = \0.4

Ukupno:

$\$2.33 + \$0.4 = \$2.73 \text{ mesečno za održavanje naše aplikacije na Lambdi.}$

- *Elastic Container Service*^{6,7,8} (ECS) je Amazonov servis koji omogućava rad sa kontejnerizovanim aplikacijama. Koristi se za aplikacije manje i srednje veličine. Servis može koristiti na dva načina: tako što se za infrastrukturu koristi EC2 ili Amazon *Fargate*. ECS kao servis ne stvara dodatni trošak, već se naplaćuje isključivo usluga EC2 ili *Fargate* servisa:
 - EC2: Ovaj način korišćenja ECS servisa stvara trošak EC2 servisa, koji zavisi od regiona, vrste mašine i količine podataka koji izađu iz mašine. Takođe, ako je potrebno, mašinama možemo dodati veće perzistentno skladište pomoću EBS servisa,
 - *Fargate*: Troškovi zavise od virtuelnih procesora, operativnog sistema, procesorske arhitekture i veličine perzistentne memorije za skladištenje. *Fargate* podržava on demand i spot načine plaćanja,
 - Cena privremenog skladišta: Plaća po satu. Prvih 20 GB je besplatno. Kada kontejner prestane sa radom, svi podaci u skladištu se brišu.
 - Kao i kod ostalih *computing* servisa, prenos podataka predstavlja trošak.

- *Elastic Kubernetes Service*^{6,7,8} (EKS) je servis koji omogućava orkestraciju kontejnerizovanih aplikacija pomoću Kubernetes alata. EKS pruža lako upravljanje i nadgledanje klastera, upravljanje kontrolnim čvorovima kao i dodavanje radnih jedinica. Slično ECS servisu, EKS se može koristiti na dva načina:

- Kreiranjem i korišćenjem EC2 instanci,
- Korišćenjem Amazon Fargate servisa.

Osim troškova koje ova dva servisa generišu, Amazon naplaćuje \$0.1 po satu po klasteru koji se koristi. Uz pomoć namespace karakteristika Kubernetesa, moguće je postaviti više aplikacija u jedan klaster. Takođe, ne treba zaboraviti ni troškove koje stvara transport podataka. EKS se koristi kada su aplikacije zahtevnije i komplikovanije. Kubernetes je alat otvorenog koda i može se koristiti na bilo kojoj infrastrukturi, dok je ECS specifičan Amazonov servis. Prilikom razvoja aplikacija i odabira rešenja predlaže se da se koristi EKS.

- *Amazon Lightsail*^{6,7,8} omogućava korisnicima brzo pokretanje i upravljanje aplikacijama bez potrebe naprednog poznavanja AWS platforme. Usluga je posebno dizajnirana za korisnike koji žele jednostavnost i olakšan pristup infrastrukturi. Koristi se za postavljanje nekih jednostavnih aplikacija koje nemaju veliki promet, veb sajtove ili testna okruženja. Korišćenje ovog servisa uključuje sledeće troškove:

- Vrsta virtuelnog servera: U zavisnosti od operativnog sistema i specifikacija hardvera, cene servera mogu ići od \$3.5 do \$160 na mesečnom nivou,
- Kontejner servisi: Ako je našim aplikacijama ili sajtovima potreban neki dodatak koji je kontejnerizovan onda se plaća dodatna cena za mašine na kojima se ti kontejneri nalaze,
- Objektno skladište: Ukoliko je potrebno skladištenje podataka kojima se ne pristupa često možemo koristiti ovaj dodatak,
- Baze podataka: Prave se nove virtuelne mašine na kojima se nalaze baze podataka. Možemo uključiti i plan visoke dostupnosti koji pravi dve instance naše baze,
- Mreža za isporuku sadržaja: Ako želimo da korisnici naše aplikacije ili veb sajta brže dobijaju sadržaj možemo koristiti *Content Delivery Network*. On se plaća po fiksnoj ceni, \$2.5 na 50 GB,
- *Load Balancer*: Servis koji nadgleda i distribuira konekcije na naše aplikacije. Ima fiksnu mesečnu cenu od \$18,
- Blokovsko skladište: Ako nam je potrebno skladište iz kog se često čitaju ili u koje se često upisuju podaci možemo koristiti ovaj dodatak,
- Stvaranje snimaka trenutnih stanja instanci i skladišta: Stvaranje rezervnih kopija je takođe moguće i plaća se po ceni od \$0.05 po GB,
- *Lightsail for research*: Omogućava kreiranje jakih mašina po pristupačnim cenama. One sadrže aplikacije za istraživanje kao što su *RStudio* i *Scilab*.

Troškovi prilikom korišćenja servisa za skladištenje

- *Amazon S3*^{6,7,8} je jedan od prvih servisa ove platforme. Pruža jednostavno skladištenje podataka po pristupačnoj ceni. Sadrži dosta opcija koje se mogu aktivirati i tako poboljšati rad sa skladištima. Od troškova imamo:
 - Cene skladištenja podataka: Osnovni trošak koji zavisi od količine podataka koja se čuva. Skladištenje može biti standardno (*standard*) čija pravila izvršava korisnik ili pametno

- (*intelligent*) koje održava veštačka inteligencija. Podaci se čuvaju u jednom regionu u minimum tri različite zone dostupnosti,
- Operacije sa skladištem: Svako čitanje, upisivanje, brisanje u S3 skladištu se naplaćuje. Cena se obračunava na hiljadu zahteva. Korisnici mesečno dobijaju 20000 GET i SELECT zahteva, 2000 PUT, COPY, POST, LIST zahteva besplatno,
 - Replikacija podataka: Podrazumeva repliciranje podataka u druga skladišta. Osim dodatne redundanse, koristi se i za bolje performanse kako bi korisnici pristupali skladištu koje je njima najbliže. Naplaćuje se po \$0.0033 po GB podataka koji se repliciraju. Replikacija može biti u okviru jednog regiona ili više regiona,
 - Transfer podataka: Svaki transfer sa interneta u naše skladište je besplatan, dok se sav transfer iz skladišta na internet plaća po fiksnoj ceni od \$0.09 po GB. Transfer podataka između S3 skladišta i ostalih Amazonovih servisa je besplatan ako se oni nalaze u istom regionu, a transfer između različitih servisa i regiona se plaća po \$0.02 po GB podataka.
- *Amazon Glacier*^{6,7,8} se nadovezuje na *Amazon S3* skladište podataka i predstavlja mogućnost arhiviranja podataka. Karakteristike ovog servisa su niska cena skladištenja ali sporiji i skuplji pristup podacima. Postoje tri vrste servisa i to su:
 - *Glacier Instant Retrieval*: Koristi se za podatke kojima se pristupa maksimalno jednom u tri meseca. Vreme pristupa je nepostojeće i svaki podatak koji se stavi u ovaj nivo skladišta mora tu biti minimalno devedeset dana. Koristi se za arhiviranje podataka. Cena ove usluge je \$0.004 po GB arhiviranog sadržaja. Zahtevi (čitanje, upisivanje, brisanje) se plaćaju \$0.02 na 1000 zahteva, dok se čitanje naplaćuje \$0.03 po GB,
 - *Glacier Flexible Retrieval*: Koristi se i za arhiviranje i za pravljenje rezervnih kopija kojima se pristupa najviše jednom u tri meseca. Vreme pristupa je između jednog minuta i dvanaest sati. Podaci se u ovom nivou skladišta najmanje mogu čuvati devedeset dana. Cena ove usluge je \$0.0036 po GB arhiviranog sadržaja. Zahtevi (čitanje, upisivanje, brisanje) se plaćaju \$0.03 na 1000 zahteva, dok se čitanje naplaćuje \$0.01 po GB,
 - *Glacier Deep Archive*: Koristi se za podatke kojima se pristupa maksimalno jednom godišnje. Vreme pristupa je do dvanaest časova. Zbog najdužeg pristupa, cena ovog skladišta je najmanja. Cena ove usluge je \$0.00099 po GB arhiviranog sadržaja. Zahtevi (čitanje, upisivanje, brisanje) se plaćaju \$0.05 na 1000 zahteva, dok se čitanje naplaćuje \$0.02 po GB.
 - *Amazon EBS*^{6,7,8} je servis koji pruža skladišne usluge na nivou bloka za integraciju sa EC2 instancama. Ova skladišta stvaraju se kao neobrađeni blokovski uređaji i mogu se montirati kao uređaji na instancama. Kada su povezani sa instancom, ponašaju se kao skladišni sistemi koji su nezavisni od vremenskog trajanja same mašine. Osnovne karakteristike EBS servisa su brz pristup i persistentno skladištenje podataka. Koriste se kao primarno skladište za fajl sisteme, baze podataka ili aplikacije koje zahtevaju malo kašnjenje. Postoje dve vrste ovog skladišta, HDD i SSD. Troškovi ovog servisa su:
 - Cene skladišta: Mere su u GB/mesečno. Ako se koriste manje od mesec dana, naplaćuju se isključivo ti dani,
 - Broj ulaznih i izlaznih operacija: Mogu se dodatno dodeliti u zavisnosti od naših potreba,

- Snimci trenutnih stanja: Mere se u GB/mesečno i možemo ih koristiti kako bi povratili naše skladište na to stanje.
- *AWS Storage Gateway*^{6,7,8} je servis koji omogućava integrisanje lokalnih okruženja sa skladištem u Oblaku. Ovaj servis dozvoljava kompanijama da jednostavno prošire svoje lokalno skladište, čime dobijaju skalabilnost i dugotrajnost *cloud* skladišta. Ovaj servis dolazi u različitim tipovima koji se naplaćuju na različite načine i svaki je dizajniran za različite upotrebe tako da imamo *File Gateway* (spaja lokalno skladište sa S3 skladištem), *Volume Gateway* (omogućava korišćenje virtuelnih fajl sistema kao da su lokalni) i *Tape Gateway* (omogućava skladištenje podataka na virtuelne magnetne trake uz pomoć *Amazon Glacier* servisa). Troškovi ovog servisa su skladištenje i pristup podacima,
- *AWS Backup*^{6,7,8} predstavlja servis u okviru AWS platforme koji omogućava korisnicima centralizovano upravljanje operacijama nad rezervnim kopijama za različite resurse. Ovaj servis je dizajniran da pojednostavi i automatizuje procese pravljenja rezervnih kopija podataka, pružajući korisnicima efikasno rešenje zaštite podataka u AWS platformi. U zavisnosti od servisa za koje pravimo rezervne kopije, cena ovog servisa može da varira.

Troškovi prilikom korišćenja servisa za baze podataka

- *Amazon RDS*^{6,7,8} (*Relational Database Service*) predstavlja servis koji omogućava kreiranje i održavanje relacionih baza podataka. Troškovi ovog servisa najviše zavise od vrste infrastrukture i skladišta te ih delimo na:
 - Infrastrukturu: Za infrastrukturu se najviše koriste t i m EC2 instance. Baze se mogu napraviti kao:
 - *Single DB Instance*: Ova opcija pravi jednu bazu podataka kojoj se pristupa. Predstavlja najjeftinije rešenje,
 - *Multi Availability Zone DB Instance*: Predstavlja glavnu bazu u jednoj zoni dostupnosti i drugu u drugoj zoni dostupnosti. Drugoj bazi se ne pristupa osim ako prva ne prestane sa radom i tada druga preuzima ulogu glavne baze. Predstavlja skuplje rešenje, ali pruža veći otpor na otkaze,
 - *Multi Availability Zone Cluster*: Predstavlja glavnu bazu i još dve baze u različitim zonama dostupnosti. Svi podaci se upisuju u glavnu bazu, a zatim se oni repliciraju na dve sporedne baze. Čitanje je moguće iz sve tri baze. Predstavlja najskuplje rešenje, ali ovaj pristup omogućava najbolje performanse i najbolju redundansu.
 - Skladište: U zavisnosti koji smo model baze izabrali (*Single DB instance*, *Multi Availability Zone DB Instance*, *Multi Availability Zone Cluster*) možemo izabrati sledeće vrste skladišta: *gp3* (standardno skladište sa dobrim uopštenim performansama) i *io1* (koristi se kod baza kojima se često pristupa),
 - Transport podataka iz i u bazu: Sa bazom se najčešće komunicira preko backend servisa. Ako su oni podignuti na EC2 instancama u istom regionu kao i baza, transport podataka je besplatan,
 - Rezervne kopije: Čuvanje rezervnih kopija baza podataka je neophodno i to se placa \$0.09 po GB skladišta,

- Transport rezervnih kopija: Trošak koji se stvara kada se rezervna kopija baze transportuje u S3 skladište. Cena ove usluge je \$0.011 po GB podataka.
- *Amazon DynamoDB*^{6,7,8} predstavlja Amazonovo rešenje za kreiranje i održavanje nerelacionih baza podataka.
 - Troškovi ovog servisa najviše zavise od načina korišćenja baze tako da postoje *On demand* (klasični model gde plaćamo onoliko koliko trošimo) i *Provisioned capacity* (koji podrazumeva zakupljivanje resursa na određeno vreme u napred),
 - Sledeći uticaj na trošak predstavlja klasa tabele:
 - *Standard*: Ova klasa se načešće koristi, operacije čitanja i pisanja su jeftinije dok je skladište skuplje,
 - *Standard Infrequent Accesses*: Koristi se kada se podacima ne pristupa često. Operacije čitanja i pisanja su skuplje dok je skladište jeftinije.
 - Skladište: Cena skladišta je fiksa i iznosi \$0.269 po GB,
 - Veličina podataka: Nerelacione baze čuvaju tekstualne podatke. Podaci do 1 KB zahtevaju 1 *write capacity unit* za upisivanje i 0.25 *read capacity unit* za čitanje. Svaki *WCU* i *RCU* se plaćaju.
- *Aurora*^{6,7,8} je Amazonovo rešenje za *PostgreSQL* i *MySQL* baze podataka. Osim osnovnih funkcija ovih baza, ovo rešenje pruža automatsko skaliranje, automatsko kreiranje rezervnih kopija, optimizovanu infrastrukturu za različite namene baza, visoku bezbednost i dostupnost. Troškovi ovog servisa zavise od:
 - Vrste instance: Postoje standardne i optimizovane za pisanje i čitanje. Kod standardnih se plaćaju skladište kao i operacije čitanja i pisanja, dok se kod optimizovanih za čitanje i pisanje plaća samo skladište,
 - Veličina instance: Kao i ostale baze, ovaj servis koristi modifikovane EC2 instance. Cene se kreću od \$0.055 do \$25 po satu, u zavisnosti od količine virtuelnih procesorskih jezgara i radne memorije,
 - Broj instanci: Moguće je koristiti veći broj instanci za čitanje podataka. One se mogu nalaziti u različitim regionima i zonama dostupnosti. Upisivanje se vrši samo u glavnu bazu. Svaka instanca se plaća posebno,
 - *Aurora Global*: Ova funkcija omogućava replikaciju baza u druge regione. Cena replikacije podataka je \$0.21 po milionu operacija pisanja,
 - Rezervne kopije: Skladištenje rezervnih kopija se plaća po GB podataka, a cena iznosi \$0.02.
- *Amazon Redshift*^{6,7,8} je servis koji omogućava rad sa velikom količinom podataka. Ima relacionu arhitekturu, a zbog svojih specifikacija se koristi u specifičnim slučajevima kao što je mašinsko učenje koje zahteva analiziranje i procesiranje više stotina terabajta podataka. Trošak ovog servisa zavisi od:
 - Arhitekture infrastrukture: Može biti klasična, u kojoj mi podešavamo sve parametre i serverless gde Amazon radi sve za nas,
 - Veličine instanci: Ove baze zahtevaju mašine sa boljom procesorskom moći. Cene se kreću od \$0.285 do \$13.675 po satu,
 - Skladišta: Skladište se plaća po fiksnoj ceni od \$0.024 po GB,

- Skladištenje rezervnih kopija: Plaća se po fiksnoj ceni od \$0.023 po GB,
 - Transfer podataka: Transfer podataka između S3 skladišta i Amazon Redshift servisa je besplatan. Svi ostali transferi se naplaćuju po ceni od \$0.02 po GB,
 - Procesiranje podataka: Procesiranje podataka se naplaćuje \$5 po TB podataka.
- Amazon DocumentDB^{6,7,8} predstavlja zamenu za MongoDB bazu podataka. Obe baze su u potpunosti kompatibilne i nema potrebe menjati kod aplikacije kako bi se ona prilagodila ovom servisu. Od troškova ovde imamo:
 - Veličine instanci: Biraju se mašine u zavisnosti od vCPU i RAM memorije,
 - Skladište: Cena skladišta po GB je \$0.10. Ono se dinamično dodeljuje bazi podataka. Ako je baza optimizovana za čitanje i upisivanje, cena je \$0.30,
 - Čitanje i upisivanje: Cena je \$0.20 za milion zahteva. Kod baza optimizovanih za čitanje i upisivanje, ovaj trošak ne postoji,
 - Rezervne kopije: Kao i kod ostalih baza, imamo jednu rezervnu kopiju za našu bazu besplatno, dok se ostale plaćaju po ceni od \$0.021 po GB,
 - Transfer: Isti kao i kod ostalih baza podataka.

Troškovi prilikom korišćenja mrežnih servisa

- Amazon Virtual Private Cloud^{6,7,8} (VPC) je servis koji korisniku omogućava kreiranje privatnog okruženja gde on može koristiti sve svoje resurse. Osnovna funkcija ovog servisa se ne plaća, dok se dodatne funkcije plaćaju i to su:
 - IP Address Manager (IPAM): Besplatan za jedan region. Omogućava pregled i menadžment privatnih i javnih IP adresa u VPC-u. Plaćena verzija ovog dodatka je \$0.00027 po aktivnoj adresi po satu,
 - NAT Gateway: Omogućava komunikaciju servisa iz privatne mreže sa resursima na internetu, ali sprečava komunikaciju u suprotnom smeru. Naplaćuje se po aktivnom satu kao i po GB podataka koji kroz njega prođu,
 - Preslikavanje saobraćaja: Ovaj dodatak kreira elastični mrežni interfejs koji preslikava saobraćaj koji kroz njega prolazi. Može se koristiti na drugim resursima. Cena ovog dodatka je \$0.018 po interfejsu,
 - Analiza dostupnosti: Predstavlja testiranje konekcije između dva resursa i pronalazi probleme ukoliko komunikacija između istih nije dostupna,
 - Analiza pristupa interfejsu: Ovaj dodatak pronalazi neidentifikovane pristupe resursima na Amazonovoj platformi. Cena analize svakog interfejsa je \$0.002,
 - Klijentski VPN: Omogućava udaljenim korisnicima da pristupaju resursima u privatnoj mreži. Ovde u trošak ulaze broj korisnika i na koliko sati se povezuju kao i broj tačaka na koje se ti korisnici mogu povezati,
 - Lokacijski VPN (Site-to-site VPN): Povezuje dve udaljene privatne mreže po ceni od \$0.05 po satu po konekciji,
 - Transit Gateway: Povezuje više privatnih mreža sa lokalnim kompanijskim mrežama i omogućava transfer saobraćaja kroz iste. Cena zavisi od broja konektovanih mreža i količine podataka koji se transferuju između tih mreža.

- *Cloudfront*^{6,7,8} predstavlja Amazonovo rešenje za *Content Delivery Network*. Kada sadržaj prvi put dođe na obod mreže, tu se kešira određeno vreme. Svaki sledeći put kada se taj sadržaj traži, on ne putuje kroz celu mrežu već se vraćai sa neke od obodnih lokacija. *Cloudfront* poboljšava performanse aplikacija i smanjuje cenu saobraćaja. Troškovi ovog servisa su:
 - Transfer podataka: Cena transfera je \$0.085 po GB, što je za \$0.005 jeftinije od običnog transfera.
 - HTTP i HTTPS zahtevi: Cena 10,000 HTTP zahteva je \$0.0075, a za 10,000 HTTPS zahteva plaćamo \$0.01.
 - Odzivi *CloudFront* funkcija: Postoji mogućnost da se jednostavnije *JavaScript* funkcije izvrše svaki put kada korisnik pošalje zahtev nekom linku ka našoj *CloudFront* distribuciji. Cena za milion izvršenih funkcija je \$0.1.
 - Čitanje iz *KeyValueStore* tabele: Čitanje se vrši prilikom izvršavanja *CloudFront* funkcija. Svaki milion čitanja koštaće nas \$0.03.
- *Elastic load balancer*^{6,7,8} je servis koji raspoređuje saobraćaj na različite destinacije i tako sprečava zagušenje mreže. Postoji više različitih vrsta, ali troškovi su isti: vreme tokom kojeg je ovaj servis aktivan i količina podataka koja se procesira,
- *Route 53*^{6,7,8} predstavlja Amazonov servis za rad sa DNS-om, tačnije, razrešavanje domena u IP adrese. Kao i kod ostalih servisa ovog tipa i ovde imamo sledeće troškove:
 - Zakup domena: Korisnici mogu zakupiti domene preko Amazonove platforme. Cena domena se računa odmah prilikom provere dostupnosti domena,
 - Održavanje domena i kreiranje rekorda: Ovaj dodatak se naziva *Hosted Zone* i omogućava korisnicima da prave rekorde svojih domena. Jedan domen može imati do deset hiljada rekorda. Svaki sledeći rekord se plaća \$0.0015 mesečno. Cena ovog servisa je \$0.50 mesečno za prvih dvadeset pet zona, a svaka zona preko tog broja se naplaćuje \$0.10 mesečno, do maksimalno pet stotina zona,
 - Upiti i zahtevi: Svaki put kada se naš domen razreši u IP adresu se naplaćuje. Postoje različiti načini rutiranja zahteva i razrešavanja domena i svaki se naplaćuje na milion upita,
 - Pravila za tok saobraćaja: Ovaj dodatak omogućava korisnicima da kreiraju kompleksnija pravila za rutiranje saobraćaja, na primer da se jedan domen razrešava u više adresa u zavisnosti od nekih kriterijuma. Cena jednog pravila je \$50 na mesečnom nivou,
 - *Route 53 Resolver*: Omogućava kreiranje privatnog domena koji funkcioniše isključivo u našoj virtuelnoj privatnoj mreži (VPC). Ako zahtevi dolaze iz našeg VPC-a, oni se ne naplaćuju, a ako dolaze iz naše lokalne mreže onda se naplaćuju po ceni od \$0.40 na milion zahteva do bilion, a onda po ceni od \$0.20 na milion zahteva.
- *Amazon Direct Connect*^{6,7,8} je servis koji omogućava spajanje naše lokalne mreže sa Amazonovom kako bi se zaobišao transport saobraćaja preko javnih putanja i umesto toga, koristila Amazonova mreža. Ovo doprinosi performansi i sigurnosti. Tri glavna troška ovde predstavljaju:
 - Kapacitet: Meri se u Mbps ili Gbps. Određuje propusni opseg,
 - Vreme aktivnosti servisa: Naplaćuje se po satu i postoje dva načina rada ovog servisa:
 - *Dedicated Connections*: Fizički portovi između naše i Amazonove mreže. Naplaćuje se dokle god je aktivan, čak i ako se ne koristi,
 - *Hosted Connections*: Logički portovi koje Amazonovi partneri stvaraju za nas i mi se preko njih konektujemo sa resursima na Amazonovoj platformi.

- Transfer saobraćaja: Cene transfera su drugačije od ostalih servisa tako da imamo:
 - Sav saobraćaj koji ulazi je besplatan,
 - Cena saobraćaja koji izlazi zavisi od toga gde se nalazi lokalna mreža koju povezujemo sa Amazonovom mrežom kao i port na koji se povezujemo.

VI. OPTIMIZACIJA TROŠKOVA

Generalna optimizacija troškova

Amazon pruža veliki broj usluga, ali se one i naplaćuju. Zbog “*pay as you go*” modela, plaćamo isključivo ono što koristimo, tako da tačno možemo izračunati koliko nam je resursa potrebno i tako što bolje optimizovati trošak. Neki od načina optimizacije su^{9,10}:

- Rezervacija resursa^{6,7}: Kada znamo koliko resursa će nam biti potrebno za neki servis (na primer za bazu podataka) možemo rezervisati Amazonove resurse na jednu ili tri godine i tako smanjiti trošak čak i do trideset odsto za plan od jedne godine i do pedeset odsto za plan od tri godine,
- Pametno upravljanje resursima^{6,7}: Sve češća mikroservisna arhitektura aplikacija zahteva distribuiranu infrastrukturu. Za razliku od tradicionalnih načina postavljanja i pokretanja aplikacija, kontejnerizacijom smanjujemo količinu potrebnih resursa. Prilikom korišćenja *Elastic Kubernetes Service* alata, mi pravimo infrastrukturu na kojoj će se pokretati naši kontejneri. Prilikom posmatranja cena EC2 instanca možemo zaključiti sledeće:
 - Cena t3.small instance je duplo jeftinija od t3.medium instance i ima samo dva GB RAM memorije manje. To znači da bi mi za istu cenu mogli da imamo dve t3.small instance koje će nam dati 2 vCPU-a više nego jedna t3.medium,
 - Cena t3.small instance je šesnaest puta manja od t3.2xlarge što znači da za istu cenu možemo imati 32 vCPU i 32 GB RAM memorije distribuirano u odnosu na jednu t3.2xlarge mašinu koja nam daje 8 vCPU i 32 GB RAM memorije.
- Na transferu^{6,7} se može uštedeti tako što će servisi aplikacije:
 - Komunicirati međusobno preko privatnih IP adresa,
 - Nalaziti se u jednoj virtuelnoj privatnoj mreži u jednom regionu i jednoj zoni dostupnosti.

Optimizacija troškova na realnom primeru

O aplikaciji

Za primer optimizacije uzećemo aplikaciju namenjenu kolekcionarima, jedan od projekata na kojima trenutno radim. Aplikacija sadrži forumski deo, gde kolekcionari mogu da diskutuju i prodavnicu gde korisnici mogu da prodaju i kupuju kolekcionarske predmete. Aplikacija je namenjena da se koristi na teritoriji Srbije, sa mogućim proširenjem u druge države u budućnosti. Aplikacija sadrži tri mikroservisa i dva baze podataka (PostgreSQL i MongoDB). Kroz testiranje, zaključili smo da nam je potrebno 5 GB RAM memorije, 12 GB skladišta i 2 vCPU za mikroservise, a za baze podataka 8 GB RAM memorije, 60 GB skladišta i 4 vCPU.

Prilikom testiranja opterećenja zaključeno je sledeće:

- Da bi prosečan broj korisnika neometano koristio aplikaciju, potrebno je: 1 servis za identifikaciju, 2 API servisa, 1 servis za plaćanje, 1 instanca MongoDB baze, 1 instanca PostgreSQL baze i 1 instanca sajta,
- Da bi maksimalni broj korisnika neometano koristio aplikaciju, potrebno je: 1 servis za identifikaciju, 3 API servisa, 1 servis za plaćanje, 1 instanca MongoDB baze, 1 instanca PostgreSQL baze i 2 instance sajta,
- Resursi za opsluživanje maksimalnog broja korisnika potrebni su 30 sati mesečno.

Procenjuje se da:

- Svakog meseca, transfer podataka ka korisnicima bi iznosio 3 TB,
- Broj zahteva na dnevnom nivou ka našoj infrastrukturi bio bi 100,000,
- Broj slika koji se čuva u skladištu bi iznosio: 5 slika po oglasu i 1 slika po profilu korisnika,
- Broj oglasa u prvih mesec dana bi porastao između 10,000 i 15,000,
- Broj korisnika bi iznosio između 25,000 i 30,000.

Troškovi bez optimizacije

Svaki trošak se računa po cenama iz evropskog (Milan, Italija) regiona, a vremenski period je 720 sati (30 dana).

- Transfer podataka: Na mesečnom nivou, iz naših servisa se ka korisnicima prosleđuje 3 TB podataka. Prvih 100 GB podataka je besplatno, tako da je trošak **\$261**,
- Korišćenje javnih IPv4 adresa: Potrebno sedam adresa (pet za mašine i dve za baze podataka) i to je ukupno **\$25.2** mesečno,
- Od računarskih servisa koristimo virtuelne mašine pošto se servisi ne kontejnerizuju tako da tu imamo pet t3.micro instance i dve t3.small instance. Ukupan trošak ovde je **\$77.76**,
- Kao *Load Balancer* koristimo *Elastic Load Balancer* servis, tačnije *Application Load Balancer*. Potrebne su nam dve instance (jedan *load balancer* radi sa saobraćajem za API servis, drugi sa instancama sajta). Ukupni troškovi ovde su **\$38.68**,
- Od servisa za skladištenje, za čuvanje slika koristimo Amazon S3 i ukupan trošak za skladištenje slika je **\$3.27**,
- Kao zamenu za MongoDB bazu podataka ćemo koristiti Amazon DynamoDB. Od troškova ovde imamo skladište, rezervne kopije i zahteve, a on iznosi **\$23.5**,
- Za PostgreSQL bazu, koristićemo Amazon RDS servis. Od troškova ovde imamo računarsku moć, skladište i rezervne kopije. On iznosi **\$93.88**.

Ukupni troškovi za neoptimizovanu infrastrukturu (slika 1) iznose:

- Transfer podataka: \$261,
- IPv4 adrese: \$25.2,
- Mrežni servisi: \$38.68,
- Računarski servisi: \$77.76,
- Servisi za skladištenje: \$3.27,
- Servisi za baze podataka: \$117.38.

Ukupno: \$523.29



Slika 1: Ukupni troškovi neoptimizovane infrastrukture

Troškovi sa optimizacijom

Kao i u prethodnom primeru, računaćemo troškove za 30 dana (720 sati), ali ćemo za region uzeti Stokholm.

- Transfer podataka: Kao i u prethodnom primeru, na mesečnom nivou, iz naših servisa se ka korisnicima prosleđuje 3 TB podataka. Korišćenjem *Cloudfront* servisa, prvih 1000 GB podataka je besplatno, tako da je ukupan trošak ovde **\$170**,
- Korišćenje javnih IPv4 adresa: Umesto javnih adresa, korišćićemo *NAT Gateway*. On će omogućiti skidanje *Docker* slika i pružiće mašinama pristup internetu. Pošto su naši servisi dostupni preko *load balancer-a*, *NAT Gateway* je potreban dvadeset sati mesečno i u tom periodu će kroz njega proći saobraćaj od 30 GB. Ukupni trošak za ovaj servis je **\$2.3**,
- Od mrežnih servisa korišćićemo *Load Balancer* i *Cloudfront*:
 - *Application Load Balancer*: Automatski se kreira kada napravimo *Ingress* objekat u našem EKS (Kubernetes) servisu za naš API mikroservis. Svi mikroservisi se mogu staviti u jedan *Ingress* i tako uštedeti novac. Troškovi ovde iznose **\$19.34**,
 - *Cloudfront*: Od troškova imamo zahteve ka slikama, što ukupno iznosi **\$1**.
- Od servisa za skladištenje, kao i u prethodnom primeru, koristimo Amazon S3. Ako slike kojima se retko pristupa (profilne slike korisnika) stavimo u *infrequent access tier*, možemo uštedeti dodatno tako da je ovde trošak **\$2.48036**,
- *DynamoDB* servis takođe koristimo i ovde tako da je trošak isto **\$23.5**,
- Umesto *Amazon RDS* servisa, uz pomoć EC2 servisa ćemo sami podići PostgreSQL bazu. Ovu mašinu možemo rezervisati na tri godine i tako dodatno uštedeti novac. Osim virtualne mašine, od troškova ovde imamo i skladište baze kao i skladištenje rezervnih kopija na S3. Ukupna cena relacije baze podataka je **\$22.95**,
- Za računarske servise korišćićemo *Kubernetes*. Pošto naša aplikacija nije velika, nećemo koristiti Amazonovo rešenje (*Elastic Kubernetes Service*) već ćemo sami podići pet t3.micro instanci rezervisanih na tri godine (*Control Plane* i *Worker Nodes*), a šesta instanca će biti aktivna samo 30 sati mesečno. Ukupni troškovi su **\$26.604**.

Ukupni troškovi za optimizovanu infrastrukturu (slika 2) iznose:

- Transfer podataka: \$170,

- IPv4 adrese: \$2.3,
- Mrežni servisi: \$20.34,
- Računarski servisi: \$26.604,
- Servis za skladištenje: \$2.48036,
- Servisi za baze podataka: \$46.45.

Ukupno: \$268.17436



Slika 2: Troškovi optimizovane infrastrukture

Troškovi optimizovane infrastrukture su 47.5%

VII. ZAKLJUČAK

Izabrani region, u kojem se nalaze naši resursi, direktno utiče na sve naše troškove¹. Ako aplikacija ne zahteva minimalna kašnjenja (poput softvera za plaćanje) možemo izabrati jeftinije regione na kontinentu gde će se naša aplikacija koristiti. Rezervacijom resursa možemo steći velike popuste za infrastrukturu. Ako znamo da će nam neke mašine biti potrebne non stop, njih možemo rezervirati. Uvek treba testirati koliko resursa je potrebno našoj aplikaciji, da ne bi došlo do neiskorišćenja istih. Korišćenjem infrastruktura kao kod alata poput *Terraform* i *Ansible*, kontejnerizacijom naše aplikacije i uz pomoć automatskog skaliranja, možemo kreirati i održavati najoptimalniju infrastrukturu. Takođe, ona se lako kreira, a kasnije i briše kada nam nije potrebna. Najbitnije je detaljno istražiti i utvrditi zahteve naše aplikacije. Amazon Web Services se svakodnevno razvija i evoluiraju, postavljajući nove standarde u oblasti *cloud computing*-a. Praćenje ovog dinamičnog ekosistema je ključno za organizacije koje žele iskoristiti najnovije tehnologije i benefite koje Amazon pruža. Uprkos stalnim promenama, održavanje informisanosti i prilagođavanje novim mogućnostima omogućava organizacijama da ostanu konkurentne i iskoriste sve prednosti koje ova tehnologija može ponuditi.

LITERATURA

- [1] Infrastruktura po IBM-u <https://www.ibm.com/topics/infrastructure>
- [2] Problemi lokalne infrastrukture <https://www.zenduty.com/blog/cloud-computing-vs-traditional-it-infrastructure/>

- [3] Računarstvo u oblaku, usluge i vrste *Cloud-a*:
https://sr.wikipedia.org/wiki/Ra%C4%8Dunarstvo_u_oblaku
- [4] Istorija AWS-a <https://press.aboutamazon.com/2002/7/amazon-com-launches-web-services-developers-can-now-incorporate-amazon-com-content-and-features-into-their-own-web-sites-extends-welcome-mat-for-developers>
- [5] AWS servisi <https://docs.aws.amazon.com/>
- [6] Troškovi i optimizacija: *Mastering AWS Cost Optimization: Real-world technical and operational cost-saving best practices* by Eli Mansoor and Yari Green, 2019.
- [7] Servisi i troškovi: *AWS: The Ultimate Guide From Beginners To Advanced For The Amazon Web Services* (2020 Edition) by Theo H. King
- [8] Troškovi: <https://aws.amazon.com/pricing>

Cost optimization in Amazon Web Services cloud

Content - The topic of this paper is *Amazon Web Services* platform, as well as its services. They are grouped, depending on the type of service that they provide, into four main groups: Computing services, database services, storage services and networking services. One of the biggest challenges while using this platform is optimizing the costs that these services create. This paper goes through all the costs of these services, protocols on how to lower them and how that all works in real life scenario where optimized and unoptimized infrastructure is compared.

Keywords - AWS, Amazon Web Services, Cloud, Infrastructure, Cost, Optimization