

# O Oracle Cloud autonomnom skladištu podataka

---

**Trstenjak, Maja**

**Master's thesis / Diplomski rad**

**2021**

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **University of Rijeka / Sveučilište u Rijeci**

*Permanent link / Trajna poveznica:* <https://um.nsk.hr/um:nbn:hr:195:283559>

*Rights / Prava:* [In copyright](#)/[Zaštićeno autorskim pravom.](#)

*Download date / Datum preuzimanja:* **2024-08-09**



*Repository / Repozitorij:*

[Repository of the University of Rijeka, Faculty of Informatics and Digital Technologies - INFORI Repository](#)



Sveučilište u Rijeci – Odjel za informatiku

Komunikacijski i informacijski sustavi

Maja Trstenjak

# O Oracle autonomnom skladištu podataka

Diplomski rad

Mentor: doc. dr. sc., Danijela Jakšić

Rijeka, 20. srpnja 2021.



Rijeka, 8.6.2021.

## Zadatak za diplomski rad

Pristupnik: Maja Trstenjak

Naziv diplomskog rada: O Oracle Cloud autonomnom skladištu podataka

Naziv diplomskog rada na eng. jeziku: About Oracle Cloud Autonomous Data Warehouse

Sadržaj zadatka:

Cilj ovog diplomskog rada je napraviti pregled Oracle Cloud autonomnog skladišta podataka (Oracle ADW). Potrebno je opisati Oracle ADW platformu te kroz praktične primjere prikazati njene značajke, funkcionalnosti, sigurnosne mogućnosti i podatkovne alate. Također, potrebno je napraviti i prikazati usporedbu Oracle ADW i AWS Redshift servisa, prema odabranim kriterijima.

Mentor:

Doc. dr. sc. Danijela Jakšić

Voditeljica za diplomske radove:

Izv. prof. dr. sc. Ana Meštrović

Zadatak preuzet: 8.6.2021.

(potpis pristupnika)

## Sadržaj

Sažetak .....	4
Ključne riječi.....	4
Abstract .....	5
Keywords .....	5
1. Uvod.....	6
2. Računarstvo u oblaku.....	7
3. Trendovi o skladištenju podataka i poslovna inteligencija (BI).....	9
4. Oracle infrastruktura u oblaku (OCI).....	11
4.1. Prednosti .....	11
4.2. Značajke.....	12
4.2.1. Sigurnost .....	12
4.2.2. Oracle autonomna baza podataka.....	12
5. Oracle autonomno skladište podataka.....	13
5.1. Prednosti .....	15
5.2. Performanse .....	16
5.3. Sigurnost .....	17
5.4. Podatkovni alati .....	19
6. AWS Redshift .....	20
7. Usporedba Oracle autonomnog skladišta podataka i AWS Redshift servisa.....	21
8. Praktični rad .....	25
8.1. Skup podataka .....	25
8.2. Postavljanje .....	30
8.3. Obrada podataka .....	32
8.3.1. Učitavanje podataka .....	33
8.3.2. Operacije nad podacima.....	36
9. Analiza Oracle ADW sustava .....	41
10. Zaključak.....	43
Popis literature .....	44
Popis slika .....	46
Popis tablica .....	47
Popis grafova .....	47

## Sažetak

Ovaj rad sastoji se od teorijskog prikaza Oracle autonomnog skladišta podataka uz objašnjenje računarstva u oblaku i Oracle infrastructure u oblaku. Razrađene su značajke koje obilježuju Oracle ADW (end. Autonomous Data Warehouse), te iznesene prednosti njegova korištenja.

U nastavku su opisani podatkovni alati, nakon čega slijedi praktični dio rada kroz koji je prikazana uporaba opisanih alata za učitavanje, transformaciju, obradu i grafički prikaz rezultata.

Dodatno je dana usporedba dva značajna dobavljača u oblaku: Oracle ADW i Amazon Redshift.

U praktičnom djelu opisan je korišten skup podataka, dan je prikaz prethodno objašnjenih podatkovnih alata koje ADW nudi te su rezultati grafički prikazani. Svi koraci obrade potkrijepljeni su odgovarajućim vizualnim prikazima.

## Ključne riječi

računarstvo u oblaku, infrastruktura u oblaku, Oracle, AWS, autonomno skladište podataka, baza podataka

## Abstract

This paper consists of a theoretical overview of the Oracle Autonomous Data Warehouse with an explanation of cloud computing and Oracle cloud infrastructure. Features that characterize Oracle ADW (end. Autonomous Data Warehouse) have been explained, and the advantages of its use have been presented.

The data tools are described afterward, followed by a practical part of the work through which the use of the described tools for loading, transformation, processing, and graphical presentation of results is shown.

Additionally, a comparison of two significant cloud providers is given: Oracle ADW and Amazon Redshift.

In the practical part is described the used dataset. The previously explained data tools offered by ADW are presented, and the results are shown graphically. All processing steps are supported by corresponding visual representations.

## Keywords

cloud computing, cloud infrastructure, Oracle, AWS, autonomous data warehouse, database

## 1. Uvod

U ovome radu teorijski su objašnjeni koncepti računarstva u oblaku, infrastrukture u oblaku, te skladište podataka u oblaku.

U posljednjih nekoliko godina u svijetu informacijskih tehnologija sve više se priča o oblaku. Novi koncept računarstva postepeno mijenja način poslovanja i funkcioniranja organizacija svih razmjera.

Računarstvo u oblaku je tehnologija koja doživljava vrhunac razvoja. Kroz obrazac događaja u prošlosti može se zaključiti da se svakih desetak godina pojavi nova tehnologija koja mijenja način na koji se posluje i način na koji se koriste informacijske tehnologije. U devedesetima to je bio internet, a posljednjih deset godina su se pojavile društvene. U ovom desetljeću to je oblak.

Tvrtke s tradicionalnim pristupom lokalne pohrane podataka, gdje su sadržaji i poslovni program instalirani lokalno kod korisnika na računalnoj opremi, imaju pristup tim sadržajima samo ako imaju opremu pored sebe i ako se nalaze unutar organizacije. Dokumenti i sadržaji programa nisu dostupni za izmjenu izvan lokalne mreže, a nadogradnja programske opreme je skupa i kompleksna. Zbog rada korisnika na različitim lokacijama javila se potreba za programskim rješenjem koje omogućuje kolaborativan rad nad dokumentima u stvarnom vremenu. Računarstvo u oblaku daje rješenje takvih poslovnih problema, pogodno je za organizacije koje imaju potrebu za naprednim uslugama na postojećim infrastrukturama, a bez investiranja u novu infrastrukturu i kadrove.

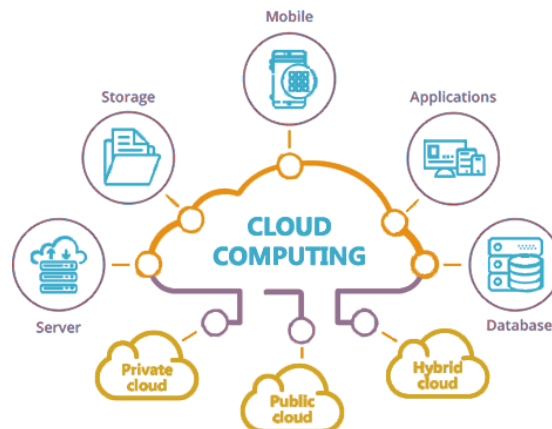
Današnji informacijski sustavi oslanjaju se na relacijske baze podataka za čuvanje i manipulaciju podacima, a takva tehnologija se primjenjuje već duži niz godina, pa shodno tome gotovo svako veće poslovanje posjeduje vlastitu bazu podataka koja često sadrži i velik broj podataka. U današnjim uvjetima sve snažnije poslovne konkurencije i informatizacije raspolaganje informacijama bitna je prednost. Složeni izračuni nad bazom podataka oduzimaju puno vremena, te se nepotrebno opterećuje baza podataka i poslovanje transakcijskog sustava. U tu svrhu, transakcijski podaci organizacija se skupljaju iz raspoloživih izvora i integriraju u jedinstveni repozitorij, skladište podataka.

Kroz napretke u tehnologiji dolazimo do skladišta podataka u oblaku. U smislu skladištenja, oblak je skup servera povezanih na internet, a na koje se mogu spremati podaci gdje se naposljetku i čuvaju. Dobro skladište u oblaku omogućava redundanciju i skalabilnost pa se skladišni prostor može širiti prema potrebi, a ako se neki server pokvari, rezervne kopije postoje na drugom serveru.

Nakon ovog uvodnog djela u radu su definirani prethodno spomenuti koncepti. Dana je definicija računarstva u oblaku, a osim oblaka, naslov teme upućuje na koncept skladišta podataka u oblaku. Glavni dio rada posvećen je Oracle skladištu podataka u oblaku, koje je obrađeno teorijski i uz praktične primjere. Uz to, kako bi se vidjele prednosti, mane i načini korištenja sustava u oblaku, odabran je AWS Redshift kao odredište podataka koje je uspoređeno s Oracle ADW. Nakon istraživanja potrebnog za ovaj rad zaključeno je kako tehnologija oblaka i skladišta u oblaku ima svjetlu budućnost i širok prostor za napredak.

## 2. Računarstvo u oblaku

Računarstvo u oblaku (eng. *cloud computing*), paradigma je informatičke tehnologije (IT) koja opisuje pružanje IT infrastrukture kao što je prostor za pohranu podataka ili aplikacijski softver kao uslugu putem Interneta. U tehničkom smislu opisuje pristup IT infrastrukture dostupnim preko računarne mreže bez potrebe instaliranja na osobno računalo. Računarstvo u oblaku omogućuje organizacijama usredotočiti se na svoje osnovne djelatnosti umjesto trošenja sredstva na računalnu infrastrukturu i održavanje, omogućuje tvrtkama da izbjegavaju ili minimiziraju troškove infrastrukture, omogućuje tvrtkama da brže iskušaju svoje aplikacije i poboljšanu upravljivost. [10]



Slika 1. Koncept računarstva u oblaku [3]

Tehnološka industrija znatno se promijenila tijekom godina, a tvrtke su iskoristile prednosti platforma u oblaku. Oblak je revolucionirao naš način poslovanja, a vodeći davatelji usluga u oblaku su:

- Oracle Corporation
- Amazon Web Services (AWS),
- Google
- Microsoft Azure.
- International Business Machines (IBM)

*Oracle Corporation* jedan je od najvećih svjetskih dobavljača na poslovnom IT tržištu i vodeći dobavljač proizvoda i usluga računalnog hardvera. Tvrtka razvija, proizvodi, prodaje i distribuira softver za baze podataka i međuprograme, aplikativni softver i hardverske sustave. Oracleovi proizvodi izgrađeni su na industrijskim standardima i projektirani su da rade zajedno ili neovisno u okviru postojeće informacijske tehnologije o kupcima, uključujući privatno i javno okruženje za računanje u oblaku.

*Oracle Corporation* osnovali su 1977. godine kao tvrtku pod nazivom *Software Development Laboratories* (SDL), Larry Ellison, Bob Miner i Ed Oates. Tvrtka je službeno preimenovana u *Oracle Corporation* 1982. godine.



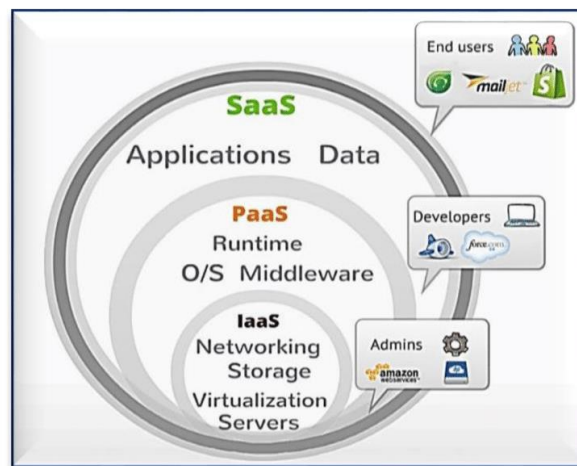
## O Oracle autonomnom skladištu podataka

Oracle poslovanje je organizirano u smjera:

- softverske sustave,
- hardverske sustave,
- usluge koje su dalje podijeljene u nekoliko operativnih segmenata

Posao softvera uključuje nove softverske licence i ažuriranja softverskih licenci te podršku proizvoda. Hardverski sustavi sastoje se od proizvoda i podrške hardverskog sustava. Usluge se sastoje od savjetovanja, usluga u oblaku i obrazovanja. Oracle također nudi alate za poslovnu inteligenciju i poslovne aplikacije.

Oracle Cloud nudi ogroman izbor rješenja za računalstvo u oblaku poslovnog razreda, uključujući PaaS, IaaS i SaaS.



Slika 2. IaaS, PaaS i SaaS [16]

Oracle Cloud zasnovan je na Javi, SQL standardima i softverskim sustavima kao što su *Exadata*, *Exalogic*, *WebLogic* i *Oracle Database* te je zasnovan je I na vrhunskim tehnologijama što ga čini bržim i učinkovitijim od konkurencije, uključujući IBM Cloud.

### 3. Trendovi o skladištenju podataka i poslovna inteligencija (BI)

Poslovna inteligencija (BI, eng. *Business intelligence*) i skladištenje podataka (DW, eng. *Data Warehouse*) dugo su bili gotovo sinonimi, te jedno bez drugog nije bilo izvedivo. Za pravodobnu analizu ogromnih povijesnih podataka potrebno je podatke organizirati, objediniti i sažeti u određeni format u skladištu podataka. Danas tvrtke imaju pristup ogromnim količinama podataka koje mogu koristiti za dobivanje poslovne inteligencije, za brže i bolje donošenje odluka. Raznolikost i složenost ovih podataka potiču potrebu za učinkovitim, isplativim načinima stvaranja analitike podataka. U svojoj osnovi poslovna inteligencija je sposobnost odgovaranja na složena pitanja o podacima i korištenja tih odgovora za donošenje poslovnih odluka. Za navedeno, potrebno je skladište podataka, koje pruža siguran način za centralizaciju i pohranu svih podataka i način za brzo pronalaženje odgovora koji su potrebni.

Tvrtke barataju velikim količinama podataka iz različitih izvora, te je nemoguće donijeti važne poslovne odluke bez BI alata koji uzimaju u obzir sve podatke. Poslovni menadžeri i analitičari podataka koriste BI alate za pronalaženje podataka kako bi saznali što se dogodilo i što se trenutno događa na poslovnom planu, te predložili radnje koje se trebaju poduzeti. BI platforme u oblaku uzimaju informacije izvedene iz podataka i prezentiraju ih u vizualnom formatu. U BI skladišta podataka služe kao okosnica za pohranu podataka. Poslovna inteligencija oslanja se na složene upite i usporedbu više skupova podataka, a učinkovito skladište podataka može ubrzati vrijeme učitavanja za pripremu i analizu podataka. Također može poboljšati sigurnost, usklađenost podataka, kapacitet pohrane, kompatibilnost integracije i dijeljenje podataka. Dok je poslovna inteligencija sustav za izvlačenje značajki povezanih uz poslovanje, skladište podataka je skladište s povijesnim i trenutnim podacima iz različitih izvora, a čine ulaz u poslovnu inteligenciju. Rezultat poslovne inteligencije su izvješća, grafovi ili dijagrami, a rezultat skladišta podataka su podaci u dimenzijskim tablicama koje poslovna inteligencija koristi. Poslovnim inteligencijom barataju menadžeri, a skladištima podataka podatkovni inženjeri i analitičari.

Skladišta podataka temelj su poslovne inteligencije koji kopira podatke iz više izvora u spremište koje može poslužiti kao jedan izvor koji je strukturiran i optimiziran za analitičku obradu i izvješćivanje. Prije nekoliko godina skladišta podataka bili su lokalni sustavi s fiksnim kapacitetom za obradu i pohranu, što je značilo da nisu mogli lako ili brzo skalirati velike obrade. Danas skladišta podataka u oblaku zamjenjuju neke naslijeđene lokalne sustave i popularan su izbor mnogih organizacija koje usvajaju nove sustave. Skladište podataka u oblaku koristi prostor i računalnu snagu koje dodjeljuje davatelj usluga oblaka, a neki primjeri su *Snowflake*, *Amazon Redshift*, *Microsoft Azure SQL DW*, *Google BigQuery* i *Oracle DW*. Navedeni servisi imaju brojne prednosti u odnosu na lokalne sustave poput brzine i skalabilnosti, uštede, sigurnosti i dostupnosti.

Inženjeri podataka, administratori baza podataka i drugi stručnjaci za podatke pomažu identificirati izvore podataka i dizajnirati skladište podataka neke organizacije. Zatim ga popunjavaju podacima pomoću ETL alata za izdvajanje, pretvaranje i učitavanje podataka iz izvora podataka. Tvrtke mogu pojednostaviti unošenje podataka korištenjem ETL alata u oblaku koji redovito replicira izvorne podatke u skladište podataka. Umjesto stručnjaka za podatke koji moraju napisati ETL kod i upravljati infrastrukturom, ETL alati automatiziraju proces povezivanja i izdvajanja podataka iz izvora i učitavanja podataka na odredište. Nakon što su podaci u skladištu podataka, mogu se

## O Oracle autonomnom skladištu podataka

analizirati bilo kojim alatom za poslovnu inteligenciju, poput *Microsoft Power BI*, *Google Data Studio*, *Tableau* ili *Metabase*.

Kao i zadnjih nekoliko godina, u budućnosti se očekuje intenzivan razvoj poslovne inteligencije, a podaci i upravljanje kvalitetom podataka bit će i dalje na prvom mjestu. Organizacije žele nadići prikupljanje što je moguće više podataka i aktivno ih koristiti za poboljšanje poslovnih odluka, a tome u prilog ide modernizacija skladišta podataka, jer su rukovanje podacima i upravljanje kvalitetom podataka posljednjih godina postali znatno složeniji. Veliki izazov dolazi s povećanjem količine podataka i broja izvora podataka, kao i uključivanjem alternativnih informacija, poput geopodataka. Pouzdani podaci i poslovna inteligencija ovise o jasnim smjericama i točno definiranim standardima za osiguravanje kvalitete, održavanja, zaštite i sigurnosti podataka stoga budućnost skladištenja podataka leži u skladištenju podataka u oblaku, zbog lakše dostupnosti podataka, boljih performansi i skalabilnosti. Kao jedno takvo rješenje u nastavku rada bit će opisani Oracle dobavljač usluge u oblaku i njegovo autonomno skladište podataka.

## 4. Oracle infrastruktura u oblaku (OCI)

Infrastruktura kao usluga (IaaS, eng. *Infrastructure as a service*) model je usluge računalstva u oblaku u kojem su računalni resursi hostani u javnom, privatnom ili hibridnom oblaku. Tvrtke mogu koristiti IaaS model za prebacivanje neke ili cijele upotrebe lokalne ili kolocirane infrastrukture podatkovnih centara u oblak, gdje je u vlasništvu i kojim upravlja davatelj usluga u oblaku. Ovi isplativi infrastrukturni elementi mogu uključivati računalni, mrežni i hardver za pohranu, kao i ostale komponente i softver. U IaaS modelu davatelj oblaka posjeduje i upravlja hardverom i softverom, a također posjeduje ili iznajmljuje podatkovni centar. Kod IaaS rješenja, unajmljuju se resursi poput računanja ili pohrane, osiguravanja po potrebi, a plaćaju se resursi koje organizacija troši. Za neke resurse, poput izračuna, plaćaju se resursi koji se koriste. Za ostale, poput skladišta, plaća se kapacitet.

Oracle infrastruktura u oblaku (OCI, eng. *Oracle Cloud Infrastructure*) duboka je i široka platforma javnih usluga u oblaku koja korisnicima omogućuje izgradnju i pokretanje širokog spektra aplikacija u skalabilnom, sigurnom, visoko dostupnom okruženju i visokih performansi.

### 4.1. Prednosti

#### 1. Autonomni servisi

OCI je ekskluzivni dom Oracle autonomne baze podataka i njezinih samopopravljivih, samooptimizirajućih autonomnih značajki. Koristeći strojno učenje za automatizaciju rutinskih zadataka, autonomna baza podataka pruža veće performanse, bolju sigurnost i poboljšanu operativnu učinkovitost te oslobađa više vremena za fokusiranje na izgradnju poslovnih aplikacija.

#### 2. Smanjenje troškova i poboljšanje performansi

OCI je pogodan za organizacije koje traže veće performanse, niže troškove i lakšu migraciju u oblak za svoje postojeće lokalne aplikacije, a uz smanjenje troškova i poboljšanje performansi računalne platforme.

#### 3. Jednostavna migracija poslovne aplikacije

Oracle Cloud je dizajniran za isporuku računalnih usluga, izolaciju mrežnog prometa, omogućuje brzu migraciju i kraće vrijeme za inovacije.

#### 4. Podrška za hibridne arhitekture

Moguća primjena programa u oblaku i baze podataka na široki izbor javnih regija do rubnih uređaja. Uz javnu regiju u oblaku, postoji mogućnost korištenja privatne namjenske regije u podatkovnim centrima.

## 4.2. Značajke

Uz prethodno opisane prednosti Oracle infrastrukture u oblaku, u nastavku je dan prikaz dviju važnih značajki: sigurnosti i autonomne baze podataka.

### 4.2.1. Sigurnost

Općenito, nijedna rasprava o infrastrukturi u oblaku ne može se dogoditi bez spominjanja sigurnosti. Oracle javni oblak druge generacije ima sigurnost kao presudan temelj, uključujući visoku izolaciju korisnika, siguran hardver domaćina, zadano šifriranje i sofisticiranu zaštitu podataka. Složenost je uzrok većine sigurnosnih problema u oblaku što dovodi do pogrešno konfiguriranih resursa i nesigurnih aktivnosti koje predstavljaju dvije različite površine napada. Oracle je iznio dvije nove ponude usluga *Oracle Maximum Security Zones* i *Oracle Cloud Guard* koje zajedno rade na smanjenju sigurnosnog rizika u oblaku, a sve bez dodatnih troškova. *Oracle Maximum Security Zones* omogućuje kupcima da automatski postave i izvrše sigurnosne politike u određenim odjeljcima za oblak koji zahtijevaju strogo sigurnosno držanje za osjetljivije podatke. Kada se stvaraju i ažuriraju resursi u sigurnosnoj zoni, *Oracle Cloud Infrastructure* provjerava operacije prema popisu definiranih pravila. *Oracle Cloud Guard* kontinuirano nadgleda konfiguracije i aktivnosti radi prepoznavanja prijetnji i automatski ih otklanja. Uz to, radi na automatskoj analizi podataka, automatskom otkrivanju prijetnji i pogrešnih konfiguracija i otklanjanju sigurnosnih prijetnji bez potrebe za ljudskim nadzorom. Proaktivno otkriva i zaustavlja nepravilne aktivnosti koje identificira, automatski isključujući zlonamjernu instancu i proaktivno ukidajući korisnička dopuštenja kada otkrije abnormalno ponašanje korisnika. *Oracle Cloud Guard* pruža sigurnosnim administratorima okvir za otkrivanje i reagiranje u oblaku potreban za skraćivanje vremena za odgovor na sigurnosne pogrešne konfiguracije i promicanje sigurnosnih operativnih centara.

### 4.2.2. Oracle autonomna baza podataka

Možda je jedna od najintragantnijih značajki Oracle oblaka autonomna baza podataka koja osigurava baze podataka, konfigurira i prilagođava određena radna opterećenja te automatski skalira računске resurse po potrebi. Oracle pozicionira autonomnu bazu podataka, ne kao dio sloja PaaS u oblaku, već kao dio Oracle infrastrukture u oblaku. [11]

Autonomnoj bazi podataka zada se razina usluge koja se treba postići, a ona rješava ostalo. Autonomna baza podataka eliminira ljudski rad osiguranja, nadzora, sigurnosti, popravljivanja i rješavanja problema te podešavanja baze podataka. Navedeno značajno smanjuje zadatke održavanja baze podataka, smanjujući troškove i oslobađajući oskudne administrativne resurse za rad na zadacima većih značaja. Nadalje, autonomna baza podataka sigurnija je od ručno upravljane baze podataka jer se sama štiti od vanjskih i unutarnjih napada, bez čekanja dostupnog administratora. Autonomna baza podataka pouzdanija je od ručno upravljane baze podataka jer se oporavlja od bilo kakvih fizičkih kvarova bilo na nivou poslužitelja ili u podatkovnom centru te ima mogućnost premotavanja podataka u prošli trenutak kako bi se poništile korisničke pogreške.

## 5. Oracle autonomno skladište podataka

Oracleova autonomna baza podataka je izgrađena na tehnologiji Oracle Exadata hardvera u kombinaciji sa softverom kao što je *Pluggable Database (PDB)*, koji izolira svakog klijenta u vlastito susjedstvo podataka ograničavajući im mogućnost gledanja izvan vlastitog PDB -a. Svi oni rade na stvarnim aplikacijskim klasterima (eng. *Real Application Clusters*, *RAC*) u kontejneru baze podataka u jednom od Oracleovih podatkovnih centara u oblaku, a Oracle upravlja cijelim sustavom, bilo na dijeljenom ili namjenskom temeljnom hardveru. Vrijedno je istaknuti i da je autonomna baza podataka relacijska baza podataka, te skalira na 128 TB, ali dolazi s funkcionalnostima sustava upravljanja relacijskim bazama podataka (RDBMS), kao što su referencijalni integritet (odnosi primarnog ključa i vanjskog ključa), te poslužiteljskih kodnih objekata poput procedura, funkcija i trigera. Važno je imati na umu da autonomna baza podataka nije samostalna usluga. To je zapravo krovni izraz za kombinaciju dvije temeljne Oracle usluge: *Autonomous Warehouse* tj. autonomno skladište podataka koje je uvelike prilagođeno za radna opterećenja skladišta podataka s podacima pohranjenim u stupčastom formatu i *Autonomous Transaction Processing* baze podataka podešene za radna opterećenja tipa *Online Transaction Processing*, formata pohrane redaka. U ovome radu fokus je na spomenutom autonomnom skladištu podataka.

Oracle autonomno skladište podataka je usluga skladišta podataka u oblaku koja uklanja sve složenosti upravljanja skladištem podataka, osiguravanjem podataka i razvojem aplikacija na temelju podataka. [5] Automatizira osiguravanje, konfiguriranje, osiguravanje, podešavanje, skaliranje i izradu sigurnosnih kopija skladišta podataka. Uključuje alate za samoposluživanje podataka, transformacije podataka, poslovne modele, automatske uvide i ugrađene mogućnosti konvergirane baze podataka koje omogućuju jednostavnije upite za više vrsta podataka i analizu strojnog učenja. Oracle autonomna DW platforma temeljena je na mehanizmima strojnog učenja s ciljem da automatski podešava u realnom vremenu zbog optimizacije performansi. Bazirana je na novoj generaciji Oracle autonomnih baza podataka koje omogućuju visoku razinu stabilnosti te elastično okruženje za upravljanje procesima.



Slika 3. Oracle skladište podataka logo [14]

## O Oracle autonomnom skladištu podataka

Oracle autonomna skladišta podataka u oblaku su dizajnirana za nekoliko profila korisnika, a koji će biti opisani u ovom radu. Za tehničke korisnike poput inženjera Oracle zauzima pristup korištenja automatizacije kako bi pružio iskustvo usmjeravanja i klika, prevlačenja i ispuštanja te korisnicima nudi intuitivna rješenja. Oracle pruža samoposlužno skladište podataka. Rad s podacima ne zahtijeva razumijevanje kako SQL funkcionira za sva ostala skladišta podataka u oblaku. Ovaj intuitivni pristup zamijenio je fizičke usluge.

Oracle je razmotrio puni životni ciklus podataka te omogućava uvid nad učitavanjem podataka, obavljanjem transformacije podataka i stvaranjem poslovnog modela. Za učitavanje podataka u skladište podataka, prvotno se stvara tablica za što je potrebno poznavanje SQL-a i razumijevanje relacijske tablice. Učitavanje podataka zahtijeva ispravne formate podataka za datoteke. Oracle je takvo rješenje usmjereno na IT prilagodio širom spektru korisnika podataka. Oracle skladište podataka u oblaku nudi korisničko sučelje (UI) i uključene alate koji lako učitavaju podatke što proces čini jednostavnim za upotrebu pod uvjetom da se radi s ispravnim podacima. Nadalje, Oracle autonomno skladište podataka nudi mogućnost učitavanja podataka s bilo kojeg mjesta, s drugih oblaka, a koristeći jednostavno sučelje za povlačenje i ispuštanje. Sustav vizualizira podatke, odlučuje o tipu podataka, omogućava korisniku da kreira novu tablicu bez SQL naredbi i pruža pregled kako će podaci izgledati kada uđu u bazu podataka. Prethodno objašnjeno korisničko sučelje trebalo bi korisnicima podataka značajno olakšati manipulaciju podataka, transformacije podataka, filtriranje i operacije čišćenja podataka.

Oracleovo skladište podataka integrira mogućnosti duboke analize u bazi podataka. Algoritmi strojnog učenja (ML) proučavaju učitane podatke i pružaju prijedloge za poslovni model, dakle nudi se mogućnost poslovnog modeliranja kao opcija. Još jedna prednost kod obrade podataka je ta da s više skupova podataka postoje različite tablice, a Oracle će automatski identificirati odnos između skupova podataka pružajući pogled na razini organizacije o tome kako će alat za poslovnu analizu tražiti podatke. Uz navedene prednosti, Oracle omogućava korištenje drugih alata nezavisnih proizvođača u smislu izvoza podataka u Excel ili vizualizacije pomoću Tableau-a. Oracle autonomno skladište podataka je otvoreno i može se koristiti s podacima i alatima u drugim oblacima, uključujući AWS i Azure.

Oracle je osmislio kako olakšati znanstvenicima mogućnosti učitavanja i transformacija uz intuitivno sučelje kojim je olakšano stvaranje modela strojnog učenja, na principu učitavanja povijesnih podataka za određene ključeve i kreiranja modela koji predviđa što će korisnici dalje raditi. Oracle je napravio veliki napredak u smislu automatizacija procesa pomoću tehnologije AutoML za izgradnju modela. AutoML je dobro definirana inovacija u znanosti o podacima s preko 20 algoritama ugrađenih u bazu podataka. Algoritam AutoML bira prave algoritamske kandidate na osnovu podataka, odabire prave attribute za hranjenje algoritama i ispravlja postavke kako bi podesio algoritme. U osnovi, postavlja konkurenciju između različitih algoritama strojnog učenja kako bi se identificirali najbolji algoritmi za određeni problem. Nakon što se skupovi podataka pripreme i izgrade modeli strojnog učenja vođeni AutoML-om, postavljanje se vrši putem REST-a. Suprotno tome, u AWS Redshift ili Snowflake nisu ugrađeni ML alati, nego se koriste vanjske usluge s odvojenim naplativim cijenama i ugovorima, što dodaje složenost i sigurnosne izazove uz stalnu potrebu za premještanjem podataka kroz baze podataka.

Line of Business Developers (LOB) su treća kategorija korisnika koju je Oracle uzeo u obzir kao potencijalne korisnike Oracle autonomnog skladišta podatka nove generacije. Za takvu vrstu rada

## O Oracle autonomnom skladištu podataka

omogućeno je brzo građenje aplikacija s malo koda. LOB programeri imaju koristi od svih novih funkcija koje su prethodno opisane i nekih postojećih alata koji su dostupni, poput APEX-a. Oracle APEX usluga za razvoj aplikacija integrirano je razvojno okruženje s niskim kodom u potpunosti integrirano sa samostalnim skladištem podataka. To je okruženje za razvoj aplikacija od kraja do kraja, koje koristi tehnike niskog koda za izgradnju, primjenu i upravljanje aplikacijama. Danas u svijetu postoji preko 500.000 korisnika APEX-a.



Slika 4. Tri glavna tipa korisnika Oracle skladišta podataka [2]

### 5.1. Prednosti

- Autonomno upravljanje skladištem

Oracle autonomno skladište podataka uklanja gotovo sve ručne administrativne zadatke, automatizira uobičajene zadatke poput sigurnosne kopije, konfiguracije i zakrpe. Također je u mogućnosti kontinuirano automatizirati podešavanje i automatsko skaliranje performansi, bez zastoja, ljudske intervencije ili pretjeranog osiguranja što poslovnim timovima omogućuje rad bez pomoći IT-a.

- Cjelovito rješenje sa samoposlužnim podatkovnim alatima i analitikom

Autonomno skladište podataka jedino je cjelovito rješenje koje koristi konvergiranu bazu podataka koja pruža ugrađenu podršku za multimodelne podatke i više radnih opterećenja. Uključuje ugrađene samoposlužne alate za poboljšanje produktivnosti analitičara, znanstvenika i programera.

- Dostupnost u Oracleovom javnom oblaku ili u podatkovnim centrima

Za razliku od ostalih usluga skladišta podataka u oblaku, Oracle autonomno skladište podataka nudi tri izbora implementacije [7]:



## O Oracle autonomnom skladištu podataka

- Zajednička infrastruktura nudi sve prednosti autonomnog skladišta podataka po nižim troškovima.
  - Namjenska infrastruktura nudi izolaciju, poboljšanu predvidljivost i kontrolu kupaca nad autonomnim politikama.
  - *Cloud @ Customer* nudi autonomno skladište podataka na Exadati u korisničkim podatkovnim centrima, udovoljavajući strogim zahtjevima za suverenost i sigurnost podataka.
- 
- Sveobuhvatna zaštita podataka i privatnosti

Autonomno skladište podataka autonomno šifrira podatke u stanju mirovanja i kretanja, štiti regulirane podatke, primjenjuje sve sigurnosne zakrpe i vrši otkrivanje prijetnji. Osim toga, korisnici mogu lako koristiti *Oracle Data Safe* za provođenje, analizu korisnika i privilegija, otkrivanje i zaštitu osjetljivih podataka i reviziju aktivnosti. Autonomno skladište podataka olakšava zaštitu podataka i od stranaca i od unutarnjih članova.

### 5.2. Performanse

Autonomno skladište podataka kontinuirano nadgleda sve aspekte performansi sustava. Autonomno vrši prilagodbe kako bi se osigurala dosljedne visoke performanse, čak i kada se opterećenja, vrste upita i broj korisnika razlikuju. Prijašnja Oracle rješenja pokazala su se dobrima na testovima, posebno u slučajevima rada s manjim količinama podataka u manjim oblacima. Danas Oracle ADW svojim rješenjima predmemoriranja podataka i paralelnog izvršavanja pokazane su performanse iznad onoga što bi se moglo očekivati od tradicionalnog mehanizma baze podataka. Kada je riječ o velikim radnim opterećenjima, ADW pruža veće performanse uz smanjuje troškova svojim operativnim prednostima i opsežnim autonomnim mogućnostima.

ADW nudi skalabilnost izračuna i pohrane, ima mogućnost automatiziranih koraka za povećavanje tj smanjivanje skalabilnosti. ADW omogućuje predmemoriranje ogromnih rezultata. Autonomna baza podataka rezultira s manje indeksa da bi se dobila bolja podrška za istodobnost. Oracle savjetnik za zdravlje klastera kontinuirano nadgleda baze podataka stvarnih klastera aplikacija (RAC, eng. *Real Application Clusters*) radi prevencije problema s izvedbom i dostupnošću. Čuvar memorije značajka je okruženja Oracle RAC za nadgledanje čvorova klastera i sprječavanje problema na čvorovima uzrokovanih nedostatkom memorije. *Hang Manager* je značajka okruženja RAC-a koja autonomno i pouzdano otkriva i rješava prekide i mrtve točke, te održava resurse na raspolaganju. Upravljanje kvalitetom usluge Oracle autonomnog skladišta podataka prilagođava konfiguraciju sustava kako bi aplikacije radile na razinama performansi potrebnim za kvalitetno poslovanje. Nadgleda i isporučuje ključne pokazatelje uspješnosti, te pruža nadzornu ploču na cijelom klasteru. Podaci iz različitih izvora i različitih formata (strukturirani i nestrukturirani) mogu se obrađivati kroz ADW bez radnih opterećenja.

### 5.3. Sigurnost

Oracle nudi sigurnost autonomnog skladišta podataka s jedinstvenim centrom za sigurnost baze podataka koji identificira osjetljive podatke i maskira ih, izdaje upozorenja o rizičnim korisnicima i konfiguracijama, kontrolira kritične aktivnosti baze podataka i otkriva sumnjive pokušaje pristupa podacima čime je smanjen rizik povrede. Automatizacija je ključ sigurnosti, strojno učenje i umjetna inteligencija smanjuju lažne pozitivne i lažni negativne rezultate te ukazuju na stvarne prijetnje uz smanjenje vremena odgovora na njih.

Organizacije mogu rasporediti korisnike, aplikacije, uređaje i infrastructure sve brže. Tijekom cijele poslovne aktivnosti, sigurnosni stručnjaci imaju dva glavna cilja: ukloniti ručne procese koji su skloni pogreškama te otkriti i otkloniti prijetnje u stvarnom vremenu, a inteligentna automatizacija daje im moć da čine oboje. Automatsko otkrivanje identificira ranjivosti i prijetnje u stvarnom vremenu, omogućujući brzi odgovor radi učinkovitog ublažavanja i unutarnjih i vanjskih kibernetičkih napada. Inteligentna automatizacija također upravlja sigurnošću zakrpe, štedeći vrijeme i ljudske resurse što je presudno za ublažavanje ranjivosti nultog dana prije nego su ti resursi iskorištavani. [6] Dakako, uvijek će postojati nove sigurnosne prijetnje evolucijom IT-a, gdje se nastoje iskoristiti ranjivosti i rizici. A da bi sigurnost organizacije išla u korak s napredcima, autonomna tehnologija mora biti ključni dio poslovanja, čime se olakšava operativni rad. Dok poslovne jedinice donose vlastite IT odluke i rade vlastite prijave i dostavu, decentralizacija IT-a unutar organizacije raste, zato su budućnost IT-a aplikacije koje moraju raditi zajedno, čak i kad im je temeljna infrastruktura različita i nisu nužno zaštićene istim mehanizmima. Da bi učinkovito zaštitio poduzeće i njegovu imovinu, sigurnosni tim mora biti interni donositelj odluka i pružatelj usluga, mora biti vijeće koje vodi smjer IT-a unutar organizacije. To podrazumijeva upotrebu inteligentnih automatizacija za osiguranje imovine gdje god se nalazila.

Podaci su najcjenjenija valuta u suvremenom poslovanju i, kao što sugerira 7,9 milijardi zapisa izloženih kršenjem podataka u 2019. godini, najveća nagrada za kibernetički kriminal. [9] Stoga nije iznenađujuće što su neke organizacije još uvijek rezervirane u pogledu prenošenja odgovornosti za podatke na pružatelja usluga javnog oblaka. Ali to ne znači da ne postoje inherentni rizici za organizaciju koja upravlja vlastitim lokalnim podatkovnim centrom. Zloupotreba privilegija, pogreške u konfiguraciji i nepoznavanje pravila uobičajeni su IT izazovi, a sve se to može ublažiti upotrebom pružatelja usluga javnog oblaka koji stavlja sigurnost na prvo mjesto. Organizacije imaju mogućnost iskoristi okretnost, fleksibilnost i skalabilnost oblaka bez ugrožavanja vlastitih podataka ili podataka svojih kupaca. Zbog toga je ugrađena sigurnost u sva rješenja u oblaku na arhitektonskoj razini, osiguravajući zaštitu u cjelini i platformu koja je dizajnerski sigurna. Uz Oracle autonomni Linux i OS menadžment u Oracle infrastrukturi u oblaku, sigurnosne zakrpe OS-a automatski se primjenjuju, što pomaže smanjiti složenost i ljudske pogreške, a istovremeno donosi povećane uštede, sigurnost i dostupnost. Nije potrebna ljudska interakcija, što automatski smanjuje rizike i pomaže u povećanju produktivnosti i dostupnosti IT osoblja.

OCI koristi sveprisutni program šifriranja za šifriranje svih podataka na svim mjestima, u svakom trenutku. Usluge *Block Volumes* i *Object Storage* omogućuju šifriranje podataka u mirovanju prema zadanim postavkama pomoću algoritma *Advanced Encryption Standard* (AES) s 256-bitnom enkripcijom. Podaci u prijevozu zaštićeni su pomoću *Transport Layer Security* (TLS) 1.2

## O Oracle autonomnom skladištu podataka

verzije. Korisnici imaju koristi od stupnjevite obrambene strategije i izuzetno sigurnih operacija koje se protežu od fizičkog sloja hardvera u Oracle podatkovnim centrima do web sloja. Mnoge zaštite i kontrole dostupne u Oracle oblaku također rade s oblacima i lokalnim rješenjima trećih strana kako bi osigurali sigurnost modernih poslovnih opterećenja i podataka.

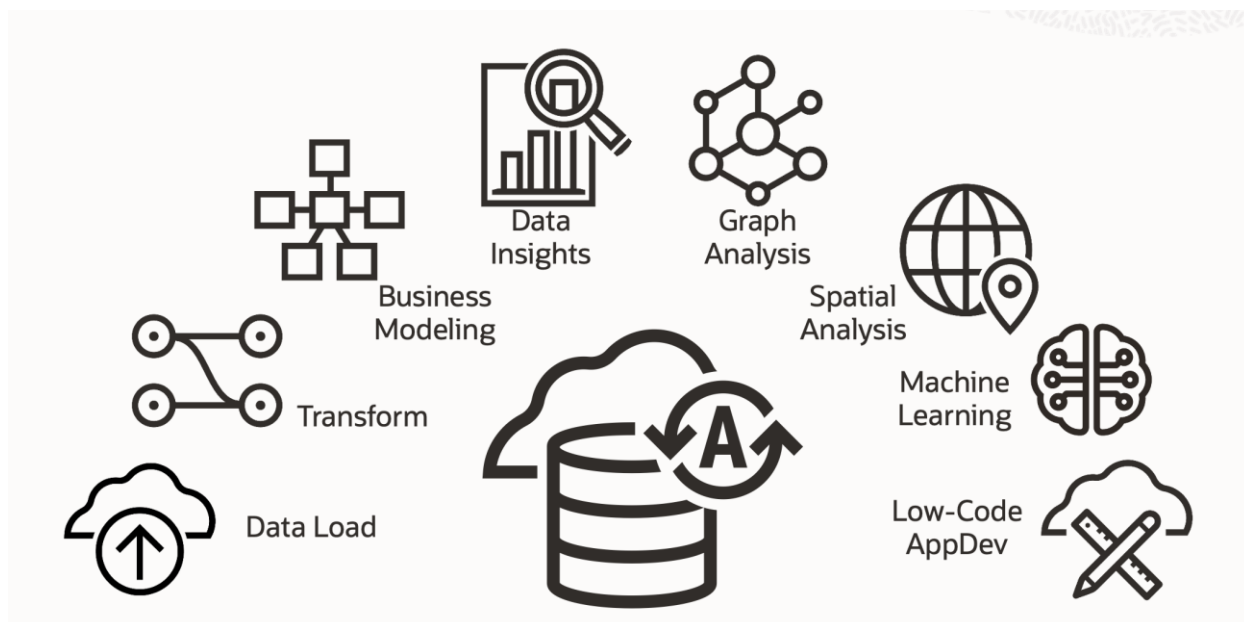
Oracle arhitektura pomaže smanjiti rizik od naprednih prijetnji i izolira podatke kako bi osigurala privatnost i sigurnost podataka. To znači da korisnici imaju koristi od:

- izolirane mrežne virtualizacije koja smanjuje rizik od napada zasnovanih na hipervizoru
- izolacije klijentovog prostora koja ograničava rizik od širenja prijetnje
- korijena povjerenja zasnovanog na hardveru koji osigurava da svaki poslužitelj ima čist firmware
- segmentacije mreže koja izolira usluge kako bi se osiguralo da se pristup kontrolira, nadgleda i pokreće strogim pravilima

Sva Oracle SaaS (eng. *Software as a service*) rješenja nude blagodati modernog oblačnog paketa. Pružaju cjelovita, agilna, sigurna i otvorena rješenja za cijelu organizaciju bez upozorenja koja dolaze s ažuriranjem i upravljanjem skupim, fizičkim lokalnim rješenjem. Oracle servis identiteta u oblaku (IDCS, eng. *Identity Cloud Service*) pruža automatizirano praćenje ponašanja u cijelom stogu i nudi sekundarnu provjeru autentičnosti poput višefaktorske provjere autentičnosti (MFA, eng. *Multi-factor authentication*). Oracle servis sigurnosti pristupa posrednika u oblaku (CASB, eng. *Cloud Access Security Broker Service*) također može analizirati ponašanja i ponašati se prema razini rizika dodijeljenoj korisniku ili aplikaciji. U slučajevima kada se utvrdi sumnjivo ponašanje, Oracle CASB može blokirati pristup korisnika visoko rizičnim uslugama, automatski upozoriti odobrene korisnike i poduzeti mjere na temelju unaprijed definiranih pravila. Oracle IDCS pomaže u ublažavanju pritiska automatskim prepoznavanjem sumnjivih transakcija. Koristeći podatke o geolokaciji, Oracle IDCS može označiti kada transakcije dolaze s neočekivanih lokacija (na primjer, kada se IP adresa korisnika ne podudara s mjestom na kojem je registriran njegov račun). Tada može primijeniti dvofaktorsku provjeru autentičnosti putem SMS-a kako bi utvrdio je li transakcija valjana. Automatskim označavanjem sumnjivih transakcija i primjenom dodatne provjere korisnika, prodavač može zaštititi svoju imovinu bez potrebe da troši značajne IT resurse.

## 5.4. Podatkovni alati

Podatkovni alati pružaju jednostavno, samoposlužno okruženje za učitavanje podataka i stavljanje na raspolaganje njihovom proširenom timu za suradnju. Analitičari poslovanja i podataka mogu lako učitati i transformirati podatke s povlačenjem-n-ispuštanjem mogućnosti, generirati poslovne modele, brzo otkriti anomalije, odstupanja i skrivene uzorke, razumjeti ovisnost podataka i utjecaj promjena.



Slika 5. Podatkovni alati dostupni kroz Oracle DB servis [15]

Slika prikazuje alate za učitavanje i transformaciju podataka, za poslovno modeliranje, za uvid u podatke, postoje i katalog te alati za strojno učenje, alati za prostorni i grafički prikaz, plus SQL Developer Web i APEX za razvoj aplikacija s niskim kodom.

U praktičnom djelu ovoga rada fokus će biti na bazama podataka i obradi podataka, te će u praktičnom dijelu rada biti prikazani alat za učitavanje podataka – *Data Load*, koji je intuitivan za korištenje. Postavlja se poveznica na lokaciju datoteke s podacima ili baze, te se nakon toga podaci transformiraju i čiste po potrebi alatom *Data Transforms*. Zatim se kroz pregledno korisničko sučelje metodom povuci – pusti kreira semantički model za što se koristi alat *Business Model*. Nakon strukturiranja podataka koristi se *Data Insight* alat gdje na temelju SQL upita dobivamo željene izlaze. Autonomno skladište podataka obrađuje podatke tražeći skrivene uzorke, anomalije i odstupanja. Za kraj rezultati će biti prikazani alatom *Graph Analysis*.

## 6. AWS Redshift

Amazon Redshift skladište je podataka u oblaku kao servis Amazon Web Services dobavljača usluga u oblaku, te je beskonačno skalabilan pomoću Amazonove oblačne infrastrukture. Redshift se može skalirati na zahtjev dodavanjem više čvorova u Redshift klaster ili stvaranjem više Redshift klastera za podršku više podataka ili bržih upita. Redshift se integrira s velikim brojem aplikacija, uključujući BI i alate za analizu, koji analitičarima omogućuju rad s podacima u Redshiftu. Redshift također radi s ETL alatima koji pomažu učitati podatke u Redshift, pripremiti ih i pretvoriti u željeno stanje.

Redshift može komunicirati s komponentama Amazon EC2 (eng. *Elastic Compute Cloud*) i S3 (eng. *Simple Storage Service*), ali njome se upravlja zasebno pomoću Redshift konzole. Na razini klastera pojavljuju se sve vrste značajki robusnosti za rješavanje rutinskih hardverskih kvarova, poput čvora ili diska; javljaju se redovite automatske sigurnosne kopije, a sigurnosne kopije na zahtjev mogu se napraviti u S3 pohranu. Moguće je dinamički promijeniti broj ili vrstu čvorova Redshifta u upotrebi. Izvorna baza podataka ostaje otvorena za upite tijekom procesa smanjivanja, no ne može se ažurirati tijekom tog procesa. Stvaranje nove instance Redshifta jednostavno je ispunjavanje nekoliko web obrazaca i čekanje da se klaster pojavi. Nakon što se klaster podigne, može se povezati s bazom podataka koristeći postavke koje su navedene pri stvaranju klastera, a zatim se prema potrebi stvaraju baze podataka, korisnici i sheme. Svi resursi mogu se razviti i definiranjem parametara kroz template dokument koji se učita u *CloudFormation* servis i kreira stog na temelju prosljeđenog dokumenta.

Amazon Redshift moćna je usluga koja je tehnologiju skladišta podataka podigla na višu razinu. Međutim, korisnici i dalje doživljavaju izazove pri postavljanju Redshifta, jer učitavanje podataka u Redshift nije trivijalno. Za cjevovode velikih razmjera, učitavanje podataka u Redshift zahtijeva postavljanje, testiranje i održavanje ETL procesa, te su potrebni stručnjaci za navedeno područje. Ažuriranja, nadogradnje i brisanja su kompleksna i moraju se obaviti pažljivo kako bi se spriječilo pogoršanje performansi upita nad podacima. Nadalje, polustrukturirane podatke je potrebno normalizirati u format relacijske baze podataka, što zahtijeva automatizaciju za velike tokove podataka. Uz to, ugniježdene strukture nisu podržane, stoga se moraju poravnati u format koji Redshift može razumjeti. Optimiziranje klastera također nije trivijalno jer postoji mnogo mogućnosti za postavljanje, te različita radna opterećenja, skupovi podataka ili čak različite vrste upita mogu zahtijevati drugačije postavljanje klastera. Tako je za visoku razinu optimizacije potrebno stalno pregledavati postavke klastera i prilagođavati broj i vrstu čvorova. I još jedna komplikacija je sigurnosno kopiranje i oporavak. Iako Amazon nudi brojne mogućnosti za izradu sigurnosne kopije skladišta podataka, te opcije su kompleksne za postavljanje i zahtijevaju nadzor.

S druge strane, Redshift nudi značajno visoku brzinu upita nad velikim skupovima, jer nudi mogućnost paralelne obrade zahtijeva. Jedna od najvećih prednosti koju pruža Redshift je fleksibilna arhitektura koja se može skalirati u nekoliko sekundi kako bi zadovoljila promjenjive zahtjeve za skladištenjem, a skalabilnost također znači uštedu, jer tvrtke nisu prisiljene trošiti novac na održavanje poslužitelja koji se ne koriste.

U nastavku rada biti će prezentirana usporedba AWS Redshift usluge i Oracle autonomnog skladišta podataka.

## 7. Usporedba Oracle autonomnog skladišta podataka i AWS Redshift servisa

Amazon Redshift je brza, potpuno upravljana usluga skladišta podataka veličine petabajta. Omogućuje jednostavnu i isplativu učinkovitu analizu podataka pomoću postojećih alata za poslovnu inteligenciju. Optimiziran je za skupove podataka u rasponu od nekoliko stotina gigabajta do petabajta. [1]

S druge strane, Oracle je sustav za upravljanje bazom podataka, RDBMS (eng. *Relational Database Management System*), koji implementira objektno orijentirane značajke kao što su korisnički definirani tipovi, nasljeđivanje i polimorfizam. Oracle baza podataka proširila je relacijski model na objektno-relacijski model, što omogućava pohranu složenih poslovnih modela u relacijsku bazu podataka.

 Oracle	Amazon AWS 
Nije potrebno podešavanje	Potrebno podešavanje
Nije potrebna administracija	Potrebna administracija

Slika 6. Glavne razlike između Oraclea i Amazon AWS-a

Slika prikazuje glavnu razliku između Oracle i AWS usluge skladišta podataka u oblaku, gdje ADW ne zahtijeva dodatnu administraciju, a Redshift zahtijeva administraciju.

Do danas postoji nekoliko varijacija portfelja Oracle autonomnih baza podataka, uključujući autonomno skladište podataka (ADW), autonomnu obradu transakcija (ATP) i autonomnu JSON bazu podataka (AJD). Za razliku od jednokratnih i ciljanih baza podataka drugih dobavljača u oblaku, Oracle Autonomous Warehouse podržava zahtjeve za više načina i za više radnih opterećenja, u okviru jednog konvergiranog mehanizma baze podataka, uključujući operativne, analitičke, JSON, grafičke, ML i blockchain dokumente. Nije potreban ETL za premještanje podataka, niti zasebne usluge za pozivanje, jer je sve dizajnirano u istoj bazi podataka. To se razlikuje od prelaska s AWS Aurore na AWS Redshift, gdje korisnici pozivaju Lambda funkciju, nakon čega slijedi Kinesis servis, u S3 segmentu, a zatim učitavaju podatke u Redshift. U Oracle Autonomous Database, mislim da se podaci o transakcijama mogu relativno lako kombinirati s analitičkim podacima. [8]

I Oracle ADW i Amazon Redshift su skladišta podataka u oblaku koja garantiraju potpuno upravljane usluge u oblaku s malo administracije, a u nastavku bit će uspoređeni po određenim mogućnostima skladišta podataka u oblaku.

Automatizacija strojnim učenjem u oblaku prva je karakteristika skladišta podataka u oblaku gdje Redshift zahtijeva ručno upravljanje i stručnjake baza podataka za analizu i dizajn te testiranje,

dok ADW ne zahtijeva stručnjake, već radi na principu “Load and Go”. Redshift se pokreće na generičkoj AWS infrastrukturi s limitiranom pouzdanošću te se temelji na PostgreSQL-u, dok ADW radi na Exadata Oracle računalnoj platformi optimiziranoj za pokretanje Oracle baza podataka. Što se tiče visine dostupnosti, oba su dobavljača visoko pouzdana sa preko 99% dostupnosti, tj. ukupno manje od 4 sata nedostupnosti na godinu. Vezano uz elastičnost i skalabilnost, ADW je samostalno skladište podataka te skaliranje, izračuni i pohrana nisu ograničeni fiksnim blokovima. U slučaju Redshifta pohrana i izračuni moraju se skalirati zajedno što uzrokuje veći trošak. U Oracle autonomnom oblaku može se osigurati do 128 CPU-a i 128 TB izravno s konzole oblaka. Redshift nije izgrađen kao baza podataka s nekoliko istodobno pokrenutih upita, a AWS preporučuje da se istodobno izvršava najviše 15 upita. Broj istodobnih korisničkih veza koje se mogu uspostaviti s klasterom je 500. Maksimalni broj tablica u Amazon Redshift-u je 9.900 za velike čvorove klastera i 20.000 za 8xlarge tipove čvorova klastera. Ograničenje uključuje privremene tablice. Oracleova baza podataka nema ograničenje broja tablica. Oracle automatski primjenjuje sva sigurnosna ažuriranja kako bi osigurao da podaci nisu ranjivi na poznate vektore napada. U Oracle ADW postoji operacija prikupljanja statistike baze podataka, i Oracle ADW i Amazon Redshift nadgledaju promjene radnog opterećenja i automatski ažuriraju statistiku u pozadini. Oracle ADW uspio je postići prikupljanje podataka uz najniže kašnjenje i postigao je najveći broj upita po satu. U smislu serijskog izvršavanja upita i protoka višekorisničkih upita, uz to dosljedno je premašio Redshift za 4 puta u tome testiranju. S druge strane, Redshift puno brže čita podatke. [4]

Objekte usluge pojednostavljuju pripremu, održavanje i optimizaciju opterećenja skladišta podataka u oblaku, te obje promoviraju ‘učitaj i povuci’ pristup analitici podataka. Korisnici samo pokreću uslugu, definiraju tablice, učitavaju podatke i zatim pokreću svoje upite. Oracle ADW prva je arhitektura koja je potpuno autonomna s ključnim atributima samoosiguranja i samopopravljanja, dok Redshift ne uključuje potpuno autonomne attribute.

### **Prednosti Oracle autonomnog skladišta podataka nad AWS Redshiftom:**

- Samo autonomno skladište podataka snižava ukupne troškove vlasništva s ugrađenim mogućnostima koje eliminiraju potrebu za višestrukim jednonamjenskim dodatnim uslugama koje zahtijevaju integraciju. ADW pruža uslugu ugrađene mogućnosti analitike I strojnog učenja koje nisu dostupne s Amazon Redshiftom. S Oracleovom uslugom korisnici mogu jednostavno stvoriti samouslužni *data mart* koristeći ugrađene mogućnosti za unos podataka, transformaciju, strojno učenje, poslovnu analitiku, grafičku analizu, geoprostornu analitiku i razvoj aplikacija. Kod AWS usluga korisnici moraju platiti i integrirati ostale servise poput Lambde, Kinesisa i S3 za premještanje podataka s Aurore u Redshift ili u Glue kako bi se izvele ETL transformacije, a pri čemu se dodatno generira i PySpark kod. Kod Oracleove usluge nema potrebe za takvim integracijama.
- Autonomno skladište podataka pruža visoku dostupnost bez zastoja za skaliranje ili održavanje. Autonomno skladište podataka omogućuje korisnicima vođenje skladišta podataka s približno 4 sata ukupnog zastoja godišnje - uključujući i planirane i neplanirane zastoje. Dok s druge strane Redshift ima 9 sati zastoja godišnje. Uz to, autonomno skladište podataka može se nadograditi bez smetnji iako usluga nastavlja raditi, dok se Redshift mora isključiti tijekom redovnog održavanja.

## O Oracle autonomnom skladištu podataka

- Autonomno skladište podataka pruža dosljedne performanse pri promjeni radnog opterećenja automatskim skaliranjem resursa i podešavanjem baze podataka. Autonomno skladište podataka automatski podešava i optimizira performanse SQL upita, neovisno o strukturi podataka ili količini podataka, dok Amazon Redshift nudi ograničen skup automatiziranih značajki za podešavanje performansi skladišta podataka. Autonomno skladište podataka automatski skalira računalne resurse kako bi povećalo performanse složenih upita u skladištu podataka i smanjilo troškove, dok Redshift nema tu mogućnost
- Autonomno skladište podataka smanjuje rizik uz sigurnost podataka i privatnost ojačanu u proizvodnji. Autonomno skladište podataka uključuje ugrađenu sigurnost s Oracle Data Safe servisom, pomažući korisnicima razumjeti osjetljivost podataka, prikriti osjetljive podatke i procijeniti rizike po podatke. Amazon Redshift nema tu razinu ugrađene funkcionalnosti, te korisnici moraju implementirati ekvivalent Oracleovih ugrađenih značajki pomoću više dodatnih alata i usluga, povećavajući složenost, potencijalne sigurnosne rizike i operativne troškove. S ADW-om šifriranje podataka o korisnicima uvijek je uključeno, a Redshift zahtijeva od korisnika da ručno uključi šifriranje koje nije automatski zadano.
- Oracle korisnicima olakšava poštivanje zahtjeva o suverenosti podataka. Autonomno skladište podataka nudi izbor implementacije u regijama širom svijeta ili u korisničkim podatkovnim centrima s rješenjima Exadata Cloud servisa tako pružajući identične arhitekture i mogućnosti lokalno i u oblaku. Time korisnici mogu lako premjestiti radna opterećenja kako bi zadovoljili promjenjive poslovne potrebe i regulatorne zahtjeve. Amazon Redshift dostupan je samo u javnom oblaku AWS-a, pa korisnici nemaju mogućnost implementacije u svojim podatkovnim centrima.

S druge strane, Redshift postiže visoke performanse pomoću masovnog paralelizma, učinkovite kompresije podataka, optimizacije upita i distribucije. Koristeći svoju arhitekturu za masovnu paralelnu obradu, Redshift može paralelizirati operacije učitavanja, sigurnosnog kopiranja i vraćanja podataka. Nadalje, upiti koji se izvršavaju distribuiraju se na više čvorova. Redshiftova stupčasta baza podataka za pohranu optimizirana je za masivne i ponavljajuće vrste podataka, dramatično smanjuje ulazno izlazne operacije na disku poboljšavajući performanse.

U nastavku je tablično prikazana usporedba Oracle i AWS pružatelja usluga u oblaku prema odabranim značajkama koje su prethodno opisane.



Tablica 1. Uporedba Oracle i AWS pružatelja usluga u oblaku prema odabranim značajkama

<b>Oracle autonomno skladište podataka</b>	<b>Mogućnosti skladišta podataka u oblaku</b>	<b>AWS Redshift</b>
Nema potrebe za ekspertima za upravljanje bazama podataka u smislu pružanja usluga, sigurnosnih kopija, osiguranja ili zakrpa.	<b>strojno učenje vođeno automatizacijom u oblaku</b>	Postoji potreba za ručno upravljanje i specijaliziranim stručnjacima za analizu i projektiranje, te testiranje prije pokretanja
Pokreće se na Exadata okosnici.	<b>dizajnirano i optimizirano za performanse skladišta podataka</b>	Radi na generičkoj AWS infrastrukturi.
Dostupnost > 99% s potpuno automatiziranim sigurnosnim kopijama.	<b>veća dostupnost</b>	Dostupnost > 99%. Ne uključuje nadogradnje, zakrpe, greške ili skaliranje.
Automatski primjenjuje sigurnosne zakrpe i nadogradnje s kompletnim izolacijama podataka.	<b>samo-osiguranost</b>	Nema izolacije operativnih korisnika od korisnika podatkovnih aplikacija.
Pojedinačno skladište podataka, skaliranje izračuna ili pohrane nije ograničeno fiksnim blokovima.	<b>elastičnost i laka skalabilnost</b>	Skladištenje i izračuni se moraju skalirati zajedno, to dovodi do potrebe za zaustavljanjem klastera, što povećava troškove.

## 8. Praktični rad

### 8.1. Skup podataka

Skup podataka korišten za potrebe ovoga rada preuzet je sa stranice <https://docs.aws.amazon.com/redshift/latest/gsg/samples/ticketdb.zip>, a naziva se *Tickit*. Ovaj primjer skupa podataka pomaže analitičarima da prate prodajne aktivnosti na izmišljenoj web stranici *TICKIT*, gdje korisnici kupuju i prodaju ulaznice za sportske događaje, predstave i koncerte putem interneta. Analitičari posebno mogu utvrditi kretanje karata tijekom vremena, stope uspješnosti prodavača i najprodavanija događanja, mjesta i godišnja doba. Također, ovi podaci mogu se koristiti za smišljanje strategija za poticanje kupaca i prodavača koji posjećuju web mjesto, privlačenje novih korisnika i poticanje oglašavanja i promocija.

*Tickit* se sastoji od sedam tablica. Detalji s opisima stupaca korištenih *Tickit* tablica prikazani su u nastavku u odgovarajućim tablicama.

- CATEGORY tablica

Tablica 2. Prikaz detalja o tablici Category

Ime stupca	Tip podataka	Opis
CATID	SMALLINT	PK, svaki redak predstavlja karakterističan tip događaja za koji je ulaznica kupljena i prodana.
CATGROUP	VARCHAR(10)	Opisni naziv grupe događaja.
CATNAME	VARCHAR(10)	Opisni naziv tipa događaja unutar grupe.
CATDESC	VARCHAR(30)	Duži opisni naziv.

- DATE tablica

Tablica 3. Prikaz detalja o tablici Date

Ime stupca	Tip podataka	Opis
DATEID	SMALLINT	PK, svaki redak predstavlja dan u kalendarskoj godini.
CALDATE	DATE	Nadnevak u formatu YYYY-MM-DD
DAY	CHAR(3)	Dan u tjednu u skraćenom obliku
WEEK	SMALLINT	Broj tjedna
MONTH	CHAR(5)	Mjesec u skraćenom obliku
QTR	CHAR(5)	Kvartal
YEAR	SMALLINT	Godina
HOLIDAY	BOOLEAN	Blagdan u US

## O Oracle autonomnom skladištu podataka

- EVENT tablica

Tablica 4. Prikaz detalja o tablici Event

Ime stupca	Tip podataka	Opis
EVENTID	INTEGER	PK, jedinstvena ID vrijednost za svaki stupac. Svaki red predstavlja zaseban događaj koji se odvija na određenom mjestu u određeno vrijeme.
VENUEID	SMALLINT	VK referenca na Venue tablicu.
CATID	SMALLINT	VK referenca na Category tablicu.
DATEID	SMALLINT	VK referenca na Date tablicu.
EVENTNAME	VARCHAR(200)	Ime događaja.
STARTTIME	TIMESTAMP	Vrijeme početka događaja.

- VENUE tablica

Tablica 5. Prikaz detalja o tablici Venue

Ime stupca	Tip podataka	Opis
VENUEID	SMALLINT	PK, svaki red predstavlja mjesto na kojem se događaji odvijaju.
VENUENAME	VARCHAR(100)	Točan naziv mjesta održavanja.
VENUECITY	VARCHAR(30)	Naziv grada.
VENUESTATE	CHAR(2)	Dvoslovna kratica države ili provincije.
VENUESEATS	INTEGER	Maksimalan broj mjesta na događaja, ako je poznat.

- USERS tablica

Tablica 6. Prikaz detalja o tablici Users

Ime stupca	Tip podataka	Opis
USERID	INTEGER	PK, svaki redak predstavlja registriranog korisnika.
USERNAME	CHAR(8)	Alfnumeričko korisničko ime od 8 znakova.
FIRSTNAME	VARCHAR(30)	Ime korisnika.
LASTNAME	VARCHAR(30)	Prezime korisnika.
CITY	VARCHAR(30)	Mjesto stanovanja korisnika.
STATE	CHAR(2)	Država
EMAIL	VARCHAR(100)	E-mail adresa korisnika.
PHONE	CHAR(14)	Telefonski broj korisnika.
LIKESPORTS, ...	BOOLEAN	Niz od 10 različitih stupaca koji identificiraju korisnikove preference.

- LISTING tablica

*Tablica 7. Prikaz detalja o tablici Listing*

Ime stupca	Tip podataka	Opis
LISTID	INTEGER	PK, jedinstvena ID vrijednost za svaki stupac. Svaki redak predstavlja popis serije ulaznica za određeni događaj.
SELLERID	INTEGER	VK referenca na Users tablicu.
EVENTID	INTEGER	VK referenca na Event tablicu.
DATEID	SMALLINT	VK referenca na Date tablicu.
NUMTICKETS	SMALLINT	Broj ulaznica dostupnih za prodaju.
PRICEPERTICKET	DECIMAL(8,2)	Fiksna cijena pojedine ulaznice.
TOTALPRICE	DECIMAL(8,2)	Ukupna cijena za seriju ulaznica (NUMTICKETS * PRICEPERTICKET)
LISTTIME	TIMESTAMP	Vrijeme objave.

- SALES tablica

*Tablica 8. Prikaz detalja o tablici Sales*

Ime stupca	Tip podataka	Opis
SALESID	INTEGER	PK, jedinstvena ID vrijednost za svaki stupac. Svaki red predstavlja prodaju jedne ili više ulaznica za određeni događaj.
LISTID	INTEGER	Foreign-key reference to the LISTING table.
SELLERID	INTEGER	VK referenca na Users tablicu. (prodavač)
BUYERID	INTEGER	VK referenca na Users tablicu. (kupac)
EVENTID	INTEGER	VK referenca na Event tablicu.
DATEID	SMALLINT	VK referenca na Date tablicu.
QTYSOLD	SMALLINT	Broj prodanih ulaznica u jednoj transakciji (max 8).
PRICEPAID	DECIMAL(8,2)	Ukupan iznos plaćen za ulaznice. Individualan iznos cijene ulaznice je PRICEPAID/QTYSOLD.
COMMISSION	DECIMAL(8,2)	Provizija od 15% koju tvrtka ubire od prodaje.
SALETIME	TIMESTAMP	Vrijeme izvršenja transakcije.

O Oracle autonomnom skladištu podataka

## RELACIJSKI MODEL TICKIT SKUPA:

Category (**catid**, catgroup, catname, catdesc)

Date (**dateid**, caldate)

Event (**eventid**, eventname, starttime, *venueid*, *catid*, *dateid*)

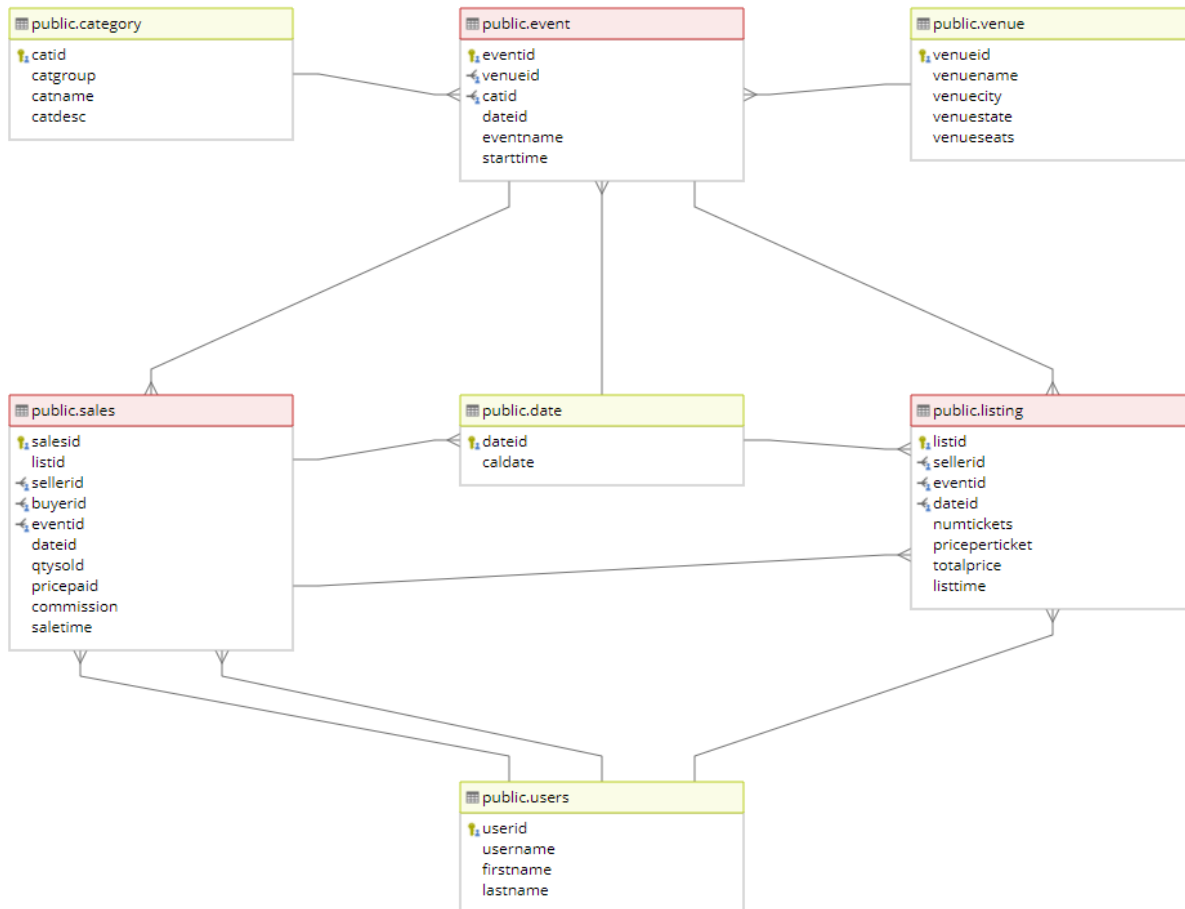
Listing (**listid**, numtickets, priceperticket, totalprice, listtime, *sellerid*, *eventid*, *dateid*)

Sales (**salesid**, qtysold, pricepaid, commission, saletime, *listid*, *sellerid*, *buyerid*, *eventid*, *dateid*)

Users (**userid**, username, firstname, lastname)

Venue (**venueid**, venue name, venuecity, venuestate, venue seats)

Na *Grafu 1.* u nastavku prikazan je model i odnos tablica sa vezama



*Graf 1. Veze između tablica Tickit skupa podataka [13]*

Na predhodnom prikazu odnosa tablica vidljivo je da *Event* tablica sadrži atribut *eventid* koji je primarni ključ, te *venueid*, *catid* i *dateid* attribute koji su vanjski ključevi.

*Category* tablica ima *catid* atribut kao primarni ključ, a *Event* tablica je povezana s njom preko tog atributa vezom *many to one*, dakle jedan događaj može se svrstati u nekoliko kategorija, ali svaka kategorija određeni događaj sadrži najviše jednom.

*Venue* tablica sadrži *venueid* atribut kao primarni ključ, te je *Event* tablica povezana s njom preko tog atributa vezom *many to one*, što znači da se jedan događaj može održavati na raznim mjestima, ali na jednom mjestu se određeni događaj održava jednom.

*Date* tablica sadrži *dateid* atribut kao primarni ključ, a *Event* tablica je povezana s njom time atributom vezom *many to one*, dakle jedan se događaj može održavati u više termina, ali se taj događaj u jednom terminu održava samo jednom. *Date* tablica je nad istim atributom povezana i sa *Sales* tablicom vezom *many to one* pa se tako za jedan termin može kupiti više karata, ali jedna karta vrijedi za samo jedan termin.

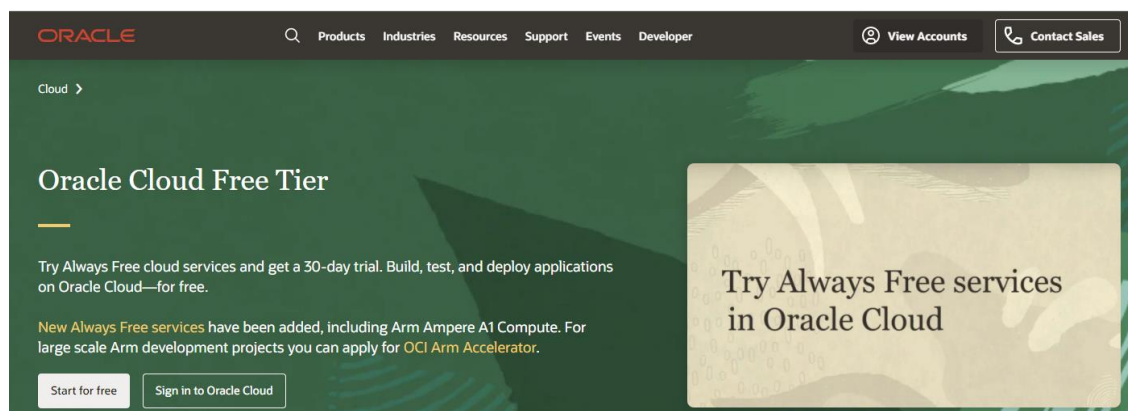
*Sales* i *Listing* tablice su povezane sa *Event* tablicom vezom *many to one* i to preko *eventid* atributa koji je u obje tablice vanjski ključ, a u *Event* tablici primarni, što znači da se za jedan događaj može kupiti više prodaja, ali jedna prodaja se odnosi samo na jedan događaj.

*Sales* tablica ima *salesid* atribut kao primarni ključ, te dodatno *sellerid* i *buyerid* kao vanjske ključeve kojima je povezana s tablicom *Users*, koja ima *userid* kao primarni ključ, vezom *many to one*. Navedeno znači da prodavač ili kupac može biti ista osoba više puta, ali za jednu prodaju je to taj korisnik samo jednom. Također *listid* atribut kojim je povezana sa *Listing* tablicom vezom *one to many*.

*Listing* tablica sadrži *listid* kao primarni ključ i *dateid* kao vanjski kojim je povezana sa *Date* tablicom vezom *many to one*, *eventid* kao vanjski ključ kojim je povezana sa *Event* tablicom vezom *many to one*, *sellerid* kao vanjski ključ kojim je povezana sa *Users* tablicom vezom *many to one*. Dakle, za jedan događaj postoji više karata, za jedan termin postoji više karata, za jednog kupca ili prodavača tj. korisnika postoji više karata, ali na jednoj karti određen je točno jedan događaj u točno jedno vrijeme održavanja, sa jednim kupcem i jednim prodavačem.

## 8.2. Postavljanje

U ovom djelu bit će opisano postavljanje računa te postavljanje infrastructure za korištenje besplatnog roka Oracle autonomnog skladišta podataka.

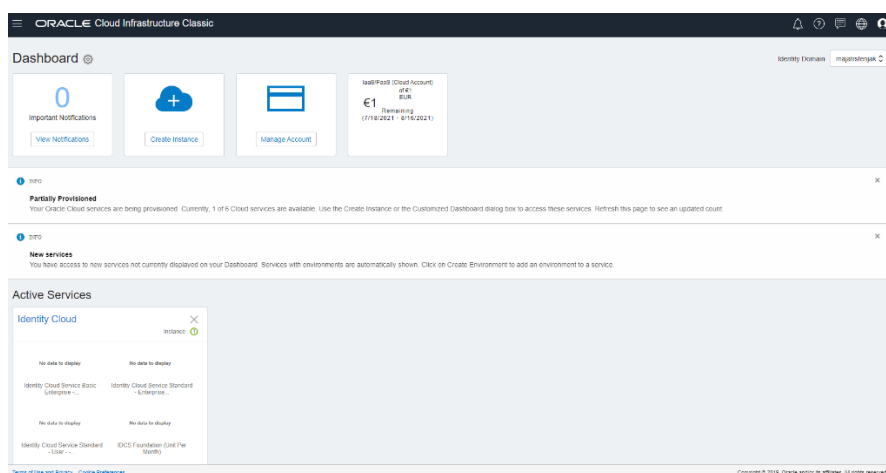


Slika 7. Pokretanje besplatnog roka korištenja Oracle servisa

Oracle servis nudi svoje alate na besplatno korištenje u probnom roku od 30 dana. Prvotno je potrebno napraviti korisnički račun, unijeti osobne podatke, e-mail adresu i kreditnu karticu, te ih je potrebno verificirati.

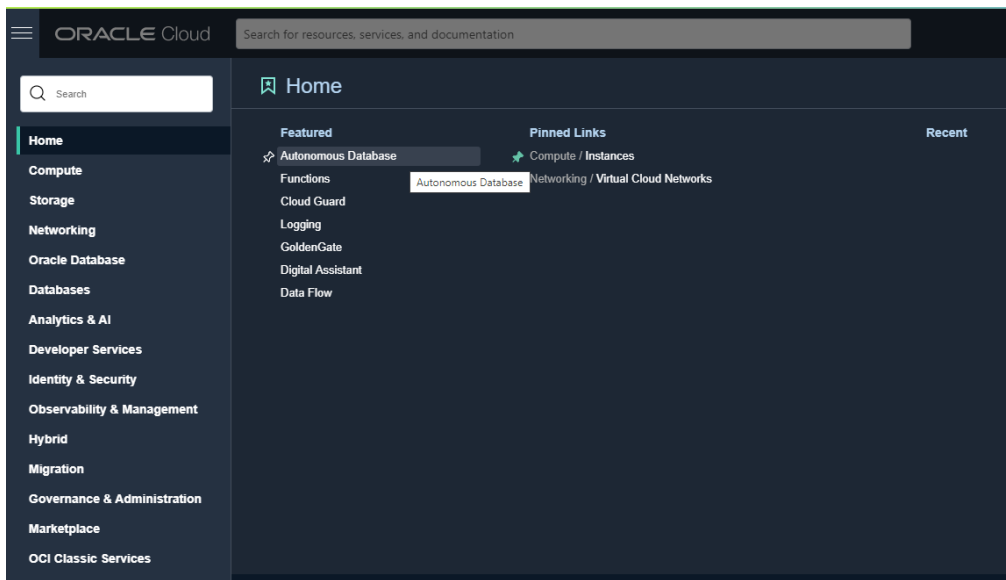
Kod prijave potrebno je odabrati regiju koja predstavlja geografsku lokaciju gdje će korišteni resursi biti kreirani. Za potrebe ovog rada odabrana je regija *Frankfurt*.

Nakon kreacije računa i verifikacije podataka prva stranica je nadzorna ploča kroz koju se vidi ponuda besplatnih servisa na raspolaganju za korištenje. Isti su prikazani na *Slici 9*.



Slika 8. Prva stranica nakon kreiranja računa na Oracle servisu je Dashboard tj nadzorna ploča

## O Oracle autonomnom skladištu podataka



Slika 9. Popis svih dostupnih Oracle servisa

Prvi servis koji se koristi za potrebe ovog rada je Autonomous Database servis kroz koji je izrađeno autonomno skladište podataka naziva *mtrstenjak\_DW\_oiri*.

### Create Autonomous Database

Provide basic information for the Autonomous Database

Compartment  
majatrstjenjak (root)

Display name  
mtrstenjak\_DW\_oiri  
A user-friendly name to help you easily identify the resource.

Database name  
mtrstenjakDW  
The name must contain only letters and numbers, starting with a letter. Maximum of 14 characters.

Choose a workload type

<b>Data Warehouse</b> Built for decision support and data warehouse workloads. Fast queries over large volumes of data. ✓	<b>Transaction Processing</b> Built for transactional workloads. High concurrency for short-running queries and transactions.	<b>JSON</b> Built for JSON-centric application development. Developer-friendly document APIs and native JSON storage.	<b>APEX</b> Built for Oracle APEX application development. Creation and deployment of low-code applications, with database included.
--	--	--	---

Choose a deployment type

<b>Shared Infrastructure</b>	Dedicated Infrastructure
------------------------------	--------------------------

Slika 10. Kreiranje skladišta podataka kroz servis baza podataka (Autonomous Database)



## O Oracle autonomnom skladištu podataka

The screenshot displays the Oracle Autonomous Database console for the instance 'mtrstenjak\_DW\_oiri'. The instance is in an 'AVAILABLE' state, as indicated by the green 'ADW' logo and the 'AVAILABLE' text below it. The console provides a comprehensive overview of the database's configuration and status.

**General Information:**

- Database Name: mtrstenjakDW
- Workload Type: Data Warehouse
- Compartment: majatrstjenjak (root)
- OCID: ...njkgq (with Show and Copy links)
- Created: Sun, Jul 18, 2021, 16:46:18 UTC
- OCPU Count: 1
- Auto Scaling: Disabled
- Storage: 20 GB
- License Type: License included
- Database Version: 19c
- Lifecycle State: Available
- Instance Type: Free (with Upgrade to Paid link)
- Mode: Read/Write (with Edit link)

**APEX Instance:**

- Instance Name: mtrstenjak\_DW\_oiri

**Infrastructure:**

- Dedicated Infrastructure: No

**Autonomous Data Guard:**

- Status: Disabled

**Backup:**

- Last Automatic Backup: No active backups exist for this database.
- Manual Backup Store: Not Configured

**Network:**

- Access Type: Allow secure access from everywhere
- Access Control List: Disabled (with Edit link)

**Maintenance:**

- Next Maintenance: Sat, Jul 24, 2021, 10:00:00 UTC - 12:00:00 UTC (with View History link)
- Customer Contacts: None (with Manage link)

**Encryption:**

- Encryption Key: Oracle-managed key

Slika 11. Kreirano skladište podataka naziva mtrstenjak\_DW\_oiri u dostupnom stanju

### 8.3. Obrada podataka

U ovom djelu bit će opisana obrada podataka kroz Oracle podatkovne alate. Nakon što je izrađeno skladište podataka na Slici 12. prikazani su podatkovni alati dostupni za korištenje.

The screenshot shows the Oracle Database Actions Launchpad, which provides a central hub for various database management and data processing tasks. The tools are organized into several categories:

- Development:**
  - SQL:** Execute queries and scripts, and create database objects.
  - REST:** Deploy REST APIs for your database.
  - CHARTS:** Visualize data from your database.
  - DATA MODELER:** Create relational diagrams for database objects.
  - JSON:** Manage your JSON Document Database.
  - APEX:** Build web applications rapidly.
- Data Tools:**
  - DATA LOAD:** Load or access data from local files or remote databases.
  - CATALOG:** Understand data dependencies and the impact of changes.
  - DATA INSIGHTS:** Discover anomalies, outliers and hidden patterns in your data.
  - BUSINESS MODELS:** Create business models for performance and analysis.
- Administration:**
  - DATABASE USERS:** Create, edit privileges and other parameters, and REST-enable database users.
- Monitoring:**
  - PERFORMANCE HUB:** View consolidated performance data for the database.

Slika 12. Dostupni podatkovni alati

### 8.3.1. Učitavanje podataka

U ovom djelu bit će opisano kako učitati podatkovni skup kako bi se obradio kroz Oracle podatkovne alate. Prvotno su kroz alat *Data Load* učitane datoteke s podacima. Slikovno su prikazani koraci na primjeru datoteke s *Event* podacima veličine 436 KB.

Na *Slici 13.* prikazan je prozor s postavkama koje se mogu urediti prema odgovarajućim vrijednostima. Za delimiter polja odabrana je *Vertikalna linija* (eng. *Pipe*) vrijednost umjesto zadane zarez (eng. *Comma*) vrijednosti. Odabrano enkodiranje je UTF-8. Zaglavlje nije učitano iz datoteke, jer prvi redak datoteke ne sadrži nazive stupaca, već su oni ručno unijeti u ovom koraku. Od podržanih tipova podataka Oracle sadrži: *varchar*, *clob*, *number*, *float*, *date*, *timestamp* i dr. U podržanim tipovima podataka nema *boolean* vrijednosti već se *boolean* polje definira kao *varchar*.

Load Data from Local File allevents\_pipe.txt (435K)

Settings | Table

File | Option: Create Table | Name: ALLEVENTS\_PIPE

Table | Properties

Error Log | Hide Properties

Encoding: 65001 - Unicode (UTF-8) | Text enclosure: None | Field delimiter: Vertical Bar

Rows to skip: 0 | Source column name:  Get from file header | Numeric column:  Convert invalid data to null

Mapping

Source column	Target column	Data Type	Format
COLUMN_1	EVENTID	NUMBER	
COLUMN_2	VENUEID	NUMBER	
COLUMN_3	CATID	NUMBER	
COLUMN_4	DATEID	NUMBER	
COLUMN_5	EVENTNAME	VARCHAR2(4000)	
COLUMN_6	STARTTIME	TIMESTAMP	RRRR-MM-DD HH24:MI:SS

Slika 13. Prozor za postavljanje parametara i tipova podataka prije njihova učitavanja u katalog

Nakon uređivanja tipova podataka, delimitera i ostalih postavki, pokreće se *Data Load Job* koji učitava podatke te su time kreirane tablice u *Catalog-u*.

## O Oracle autonomnom skladištu podataka

Entity type	Namespace	Owner	Application	Created on	Updated on
TABLE	DB	ADMIN	DATABASE	2021-07-19T18:48:47Z	2021-07-19T18:48:47Z
TABLE	DB	ADMIN	DATABASE	2021-07-19T18:47:25Z	2021-07-19T18:47:25Z
TABLE	DB	ADMIN	DATABASE	2021-07-19T18:42:14Z	2021-07-19T18:42:14Z
TABLE	DB	ADMIN	DATABASE	2021-07-19T18:41:03Z	2021-07-19T18:41:03Z
TABLE	DB	ADMIN	DATABASE	2021-07-19T18:39:43Z	2021-07-19T18:39:43Z

Slika 14. Prikaz kataloga sa nekim od kreiranih tablica

Prikaz dijela kreirane tablice nalazi se na *Slici 15*. U pregledu se između ostalog vide i statistika (eng. Statistics) i definicija podataka (eng. Data Definition).

Preview	EVENTID	VENUEID	CATID	DATEID	EVENTNAME	STARTTIME
Lineage	126	300	8	2056	The Birds (Die Vogel)	2008-08-18T20:00
Impact	127	308	8	2161	The Birds (Die Vogel)	2008-12-01T14:00
Statistics	128	303	8	2067	Rusalka	2008-08-29T14:30
Data Definition	129	306	8	2166	Salome	2008-12-06T19:00
	130	300	8	1911	La Rondine	2008-03-26T19:30
	131	300	8	1969	The Birds (Die Vogel)	2008-05-01T14:30

Slika 15. Dio Event tablice

U tabu *Statistics* prikazana je statistika po stupcima, s tipom podataka, brojem različitih vrijednosti, brojem *null* vrijednosti, prikazom minimalne i maksimalne vrijednosti i prikazom relacijskih veza. Primjer prikaza nalazi se na *Slici 16*.

## O Oracle autonomnom skladištu podataka

### ALLEVENTS\_PIPE

Preview	Table Size (in bytes):	351920	Number of Rows:	8798	Number of Columns:	6		
Lineage		Data Type	Constraint	Distinct Values	Null Values	Minimum Value	Maximum Value	
Impact	EVENTID	NUMBER		8798	0	1	8798	
Statistics	VENUEID	NUMBER		204	0	1	309	
Data Definition	CATID	NUMBER		4	0	6	9	
DATEID	NUMBER			365	0	1827	2191	
EVENTNAME	VARCHAR2(4000)			576	0	.38 Special	Zumanity	
STARTTIME	TIMESTAMP(6)			1925	0	2008-01-01 14:00:00	2008-12-31 20:00:00	

Slika 16. Prikaz statistike za Event tablicu

Definicija podataka (DDL) prikazana je na Slici 17.

### ALLEVENTS\_PIPE

Preview	1	
	2	CREATE TABLE "ADMIN"."ALLEVENTS_PIPE"
	3	( "EVENTID" NUMBER,
Lineage	4	"VENUEID" NUMBER,
	5	"CATID" NUMBER,
Impact	6	"DATEID" NUMBER,
	7	"EVENTNAME" VARCHAR2(4000) COLLATE "USING_NLS_COMP",
	8	"STARTTIME" TIMESTAMP (6)
Statistics	9	) DEFAULT COLLATION "USING_NLS_COMP" SEGMENT CREATION IMMEDIATE
	10	PCTFREE 10 PCTUSED 40 INITRANS 10 MAXTRANS 255
	11	COLUMN STORE COMPRESS FOR QUERY HIGH ROW LEVEL LOCKING LOGGING
Data Definition	12	STORAGE(INITIAL 65536 NEXT 1048576 MINEXTENTS 1 MAXEXTENTS 2147483645
	13	PCTINCREASE 0 FREELISTS 1 FREELIST GROUPS 1
	14	BUFFER_POOL DEFAULT FLASH_CACHE DEFAULT CELL_FLASH_CACHE DEFAULT)
	15	TABLESPACE "DATA" ;

Slika 17. Prikaz DDL-a Event tablice

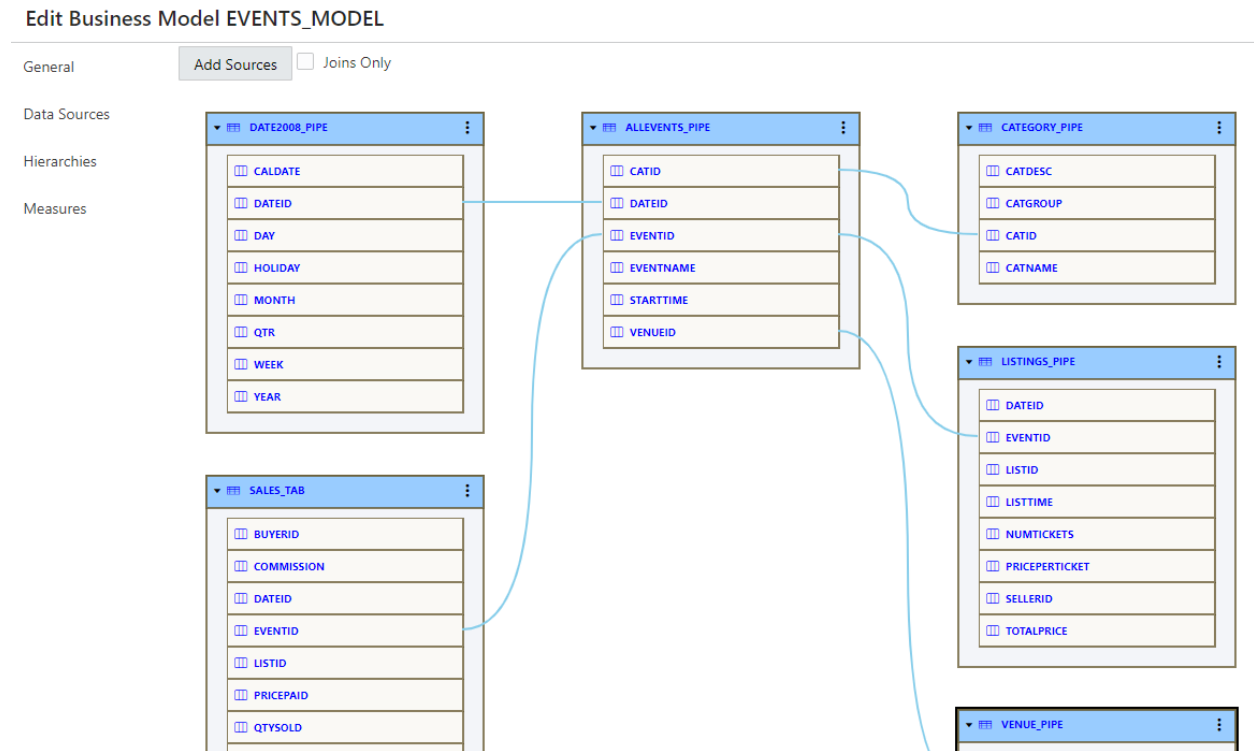
U idućem poglavlju bit će prikazane operacije nad podacima.

## O Oracle autonomnom skladištu podataka

### 8.3.2. Operacije nad podacima

U ovom djelu bit će opisana obrada podataka kroz odabrane SQL upite.

Kroz *Business Models* alat kreiran je jedan primjer modela, te su kreirane mjere na kojima će biti pokrenut *Data Insights*.



Slika 18. Kreiran model naziva *EVENTS\_MODEL\_AV*

Edit Business Model EVENTS\_MODEL

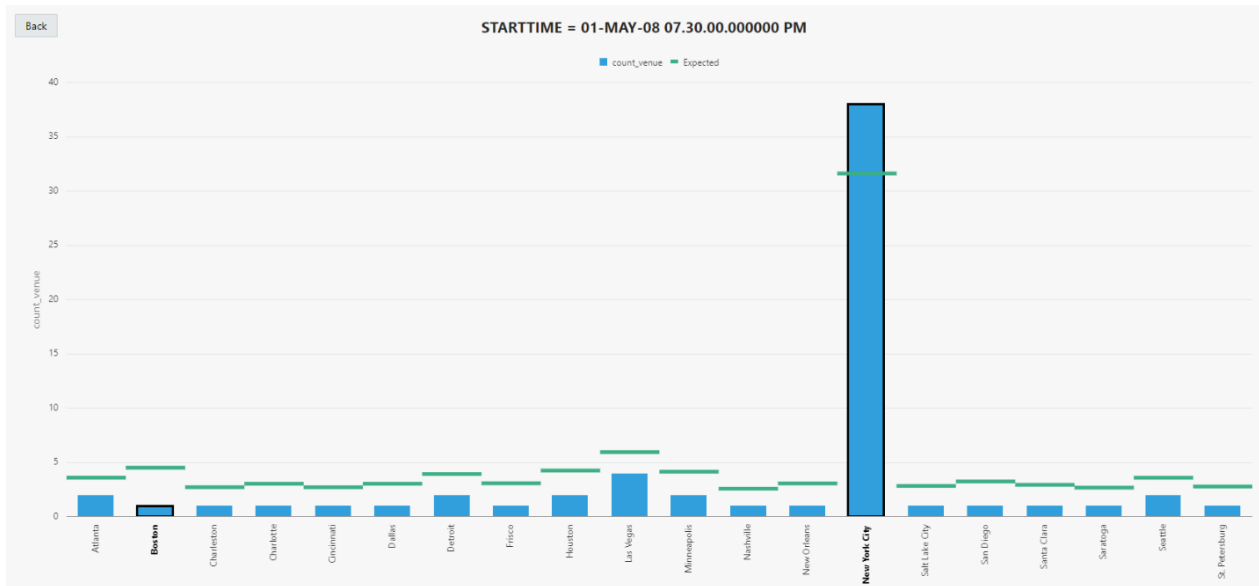
General

Measure Name	Caption	Description	Column	Expression
max_event	max_event	max value for event	EVENTNAME	MAX
count_venue	count_venue	count of events per venue	VENUEID	COUNT
sum_event	sum_event	sum of events	EVENTID	SUM

Slika 19. Kreirane mjere na temelju kojih će poslovni model biti proučavan

## O Oracle autonomnom skladištu podataka

Kroz *Data Insights* alat odabran je prethodno kreiran model naziva *EVENTS\_MODEL\_AV*, te stupac *count\_venue* tj. mjera koja je kreirana a označava broj održanih događaja po mjestu. Ovaj alat proučava podatke kako bi odredio matematički model podataka i prijavio izvanredne vrijednosti koje su mnogo veće ili niže od očekivane vrijednosti.

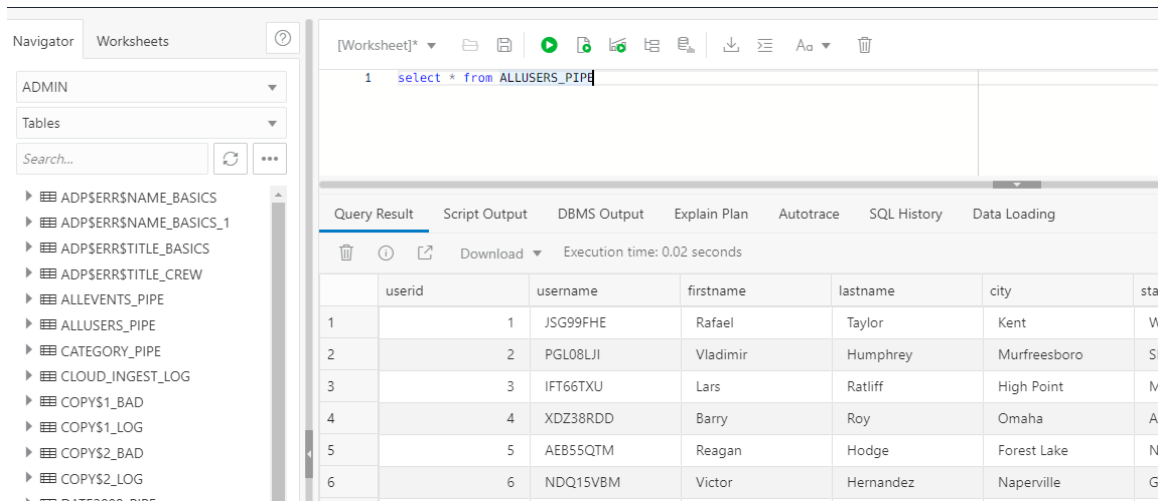


Slika 20. Grafički prikaz broja održanih događaja po mjestu uz označena značajna odstupanja

Slika 20. Prikazuje rezultat uvida dobiven *Data Insights* alatom za dan 1. svibnja 2008.godine kada je najviše događaja održano u New Yourk-u, što je prema modelu i očekivan ishod. Broj održanih događaja premašuje procijenjeni broj prema modelu. Uz New Yourk, broj održanih događaja odstupa od procijenjenog i za Boston, gdje je iznos nešto manji od predviđenog. Inače, zelenom linijom označene su očekivane vrijednosti procijenjene na temelju modela, a plavom bojom označene su stvarne vrijednosti. Crno su označena značajna odstupanja tj. anomalije.

Kroz *SQL Page* isproban je upit koji će pronaći top prodavače u San Diegu bazirano na broju prodanih ulaznica u 2008. godini. SQL stranica omogućuje unos i izvršavanje SQL i PL/SQL izraza i stvaranje objekata baze podataka. SQL izrazi uneseni u radni list (eng. Worksheet) mogu koristiti za stvaranje tablice, umetanje podataka, izradu i uređivanje okidača, odabir podataka iz tablice i spremanje tih podataka u datoteku. Dodatne značajke su isticanje sintakse i otkrivanje pogrešaka. SQL stranica sastoji se od lijevog prozora za navigaciju radnim listovima i objektima, uređivača za izvršavanje SQL izraza i izlaznog prozora za pregled rezultata. Izgled sučelja prikazan je na *Slici 21*. uz pokretanje jednostavnog upita gdje se iz *User* tablice dohvaćaju svi podaci.

## O Oracle autonomnom skladištu podataka

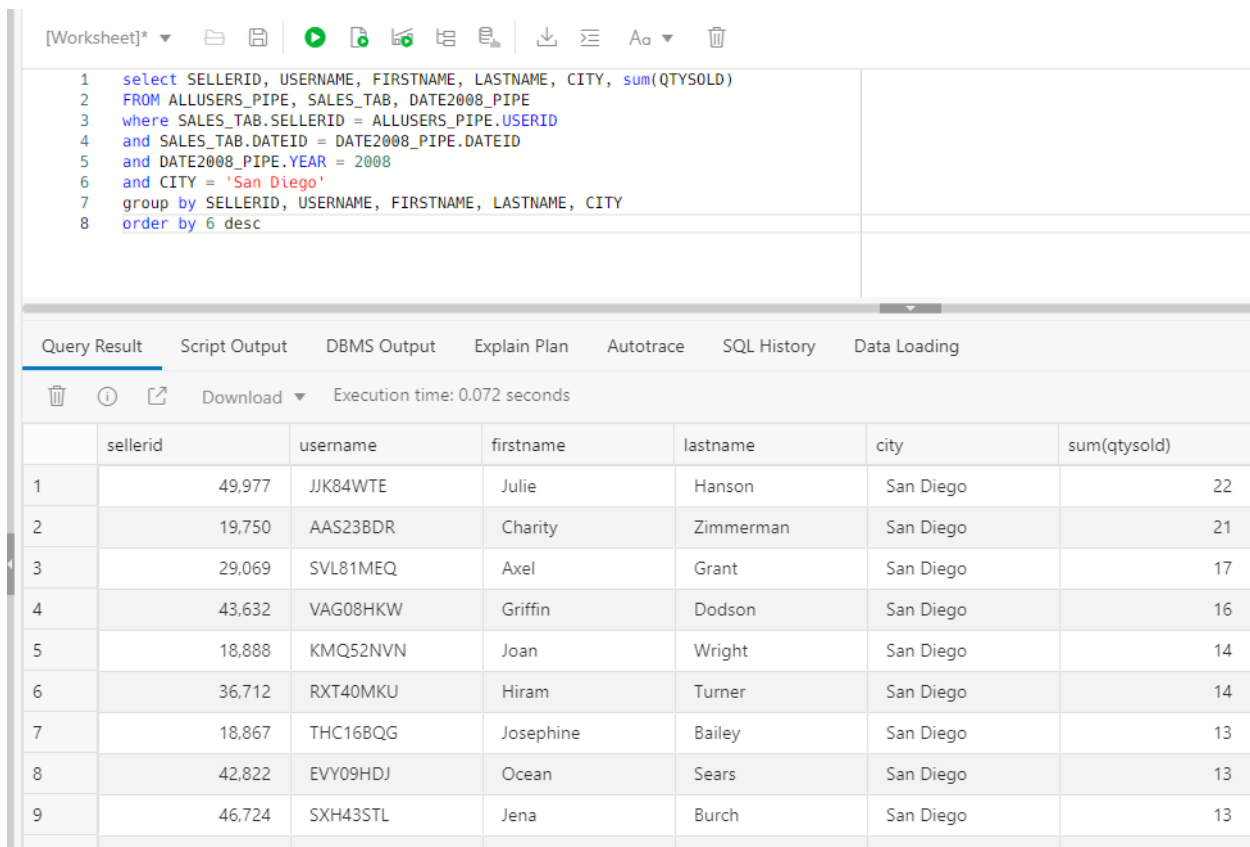


The screenshot shows the Oracle SQL Page interface. On the left is a Navigator pane with a tree view of database objects under the 'ADMIN' schema. The main area displays a query: `1 select * from ALLUSERS_PIPE`. Below the query editor, the 'Query Result' tab is active, showing a table with 6 rows and 7 columns: `userid`, `username`, `firstname`, `lastname`, `city`, and `sta`. The execution time is 0.02 seconds.

	userid	username	firstname	lastname	city	sta
1	1	JSG99FHE	Rafael	Taylor	Kent	W
2	2	PGL08LJI	Vladimir	Humphrey	Murfreesboro	S
3	3	IFT66TXU	Lars	Ratliff	High Point	M
4	4	XDZ38RDD	Barry	Roy	Omaha	A
5	5	AEB55QTM	Reagan	Hodge	Forest Lake	N
6	6	NDQ15VBM	Victor	Hernandez	Naperville	G

Slika 21. Izgled sučelja unutar SQL Page alata

Vrijeme za izvršenje upita nad *Tickit* podacima je 0.072 sekunde, a upit je pokrenut nad dvije tablice s 200000 zapisa, te nad 50000 zapisa u *Users* tablici.



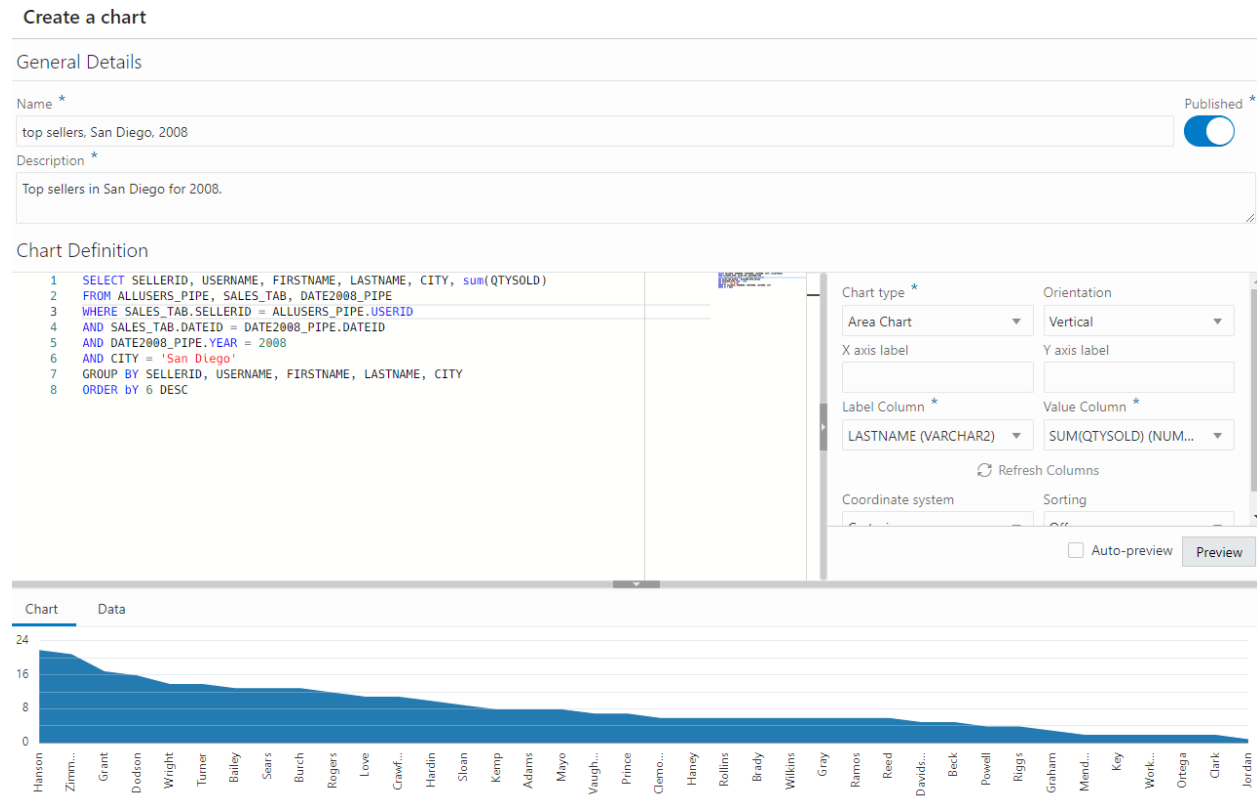
The screenshot shows a more complex query in the SQL Page interface. The query is: `1 select SELLERID, USERNAME, FIRSTNAME, LASTNAME, CITY, sum(QTYSOLD)  
2 FROM ALLUSERS_PIPE, SALES_TAB, DATE2008_PIPE  
3 where SALES_TAB.SELLERID = ALLUSERS_PIPE.USERID  
4 and SALES_TAB.DATEID = DATE2008_PIPE.DATEID  
5 and DATE2008_PIPE.YEAR = 2008  
6 and CITY = 'San Diego'  
7 group by SELLERID, USERNAME, FIRSTNAME, LASTNAME, CITY  
8 order by 6 desc`. The 'Query Result' tab shows a table with 9 rows and 7 columns: `sellerid`, `username`, `firstname`, `lastname`, `city`, and `sum(qtysold)`. The execution time is 0.072 seconds.

	sellerid	username	firstname	lastname	city	sum(qtysold)
1	49,977	JJK84WTE	Julie	Hanson	San Diego	22
2	19,750	AAS23BDR	Charity	Zimmerman	San Diego	21
3	29,069	SVL81MEQ	Axel	Grant	San Diego	17
4	43,632	VAG08HKW	Griffin	Dodson	San Diego	16
5	18,888	KMQ52NVN	Joan	Wright	San Diego	14
6	36,712	RXT40MKU	Hiram	Turner	San Diego	14
7	18,867	THC16BQG	Josephine	Bailey	San Diego	13
8	42,822	EVY09HDJ	Ocean	Sears	San Diego	13
9	46,724	SXH43STL	Jena	Burch	San Diego	13

Slika 22. Upit nad *Tickit* skupom podataka sa pripadajućim izlaznim rezultatom

## O Oracle autonomnom skladištu podataka

Za demonstraciju ovog rada isproban je i *Graph* alat koji Oracle nudi, te je za prethodni upit napravljen vizualni prikaz rezultata. U *Chart Definition* polje unosi se upit za koji želimo vizualni rezultat, te se nadalje unose vrijednosti za parametre *Chart type* koji je u ovom slučaju *Area Chart*, a za *Orientation* je odabrana vertikalna orijentacija. Na *x* osi prikazana su prezimena korisnika, a na *y* osi prikazana je suma prodanih ulaznica pod njihovim korisničkim brojem.

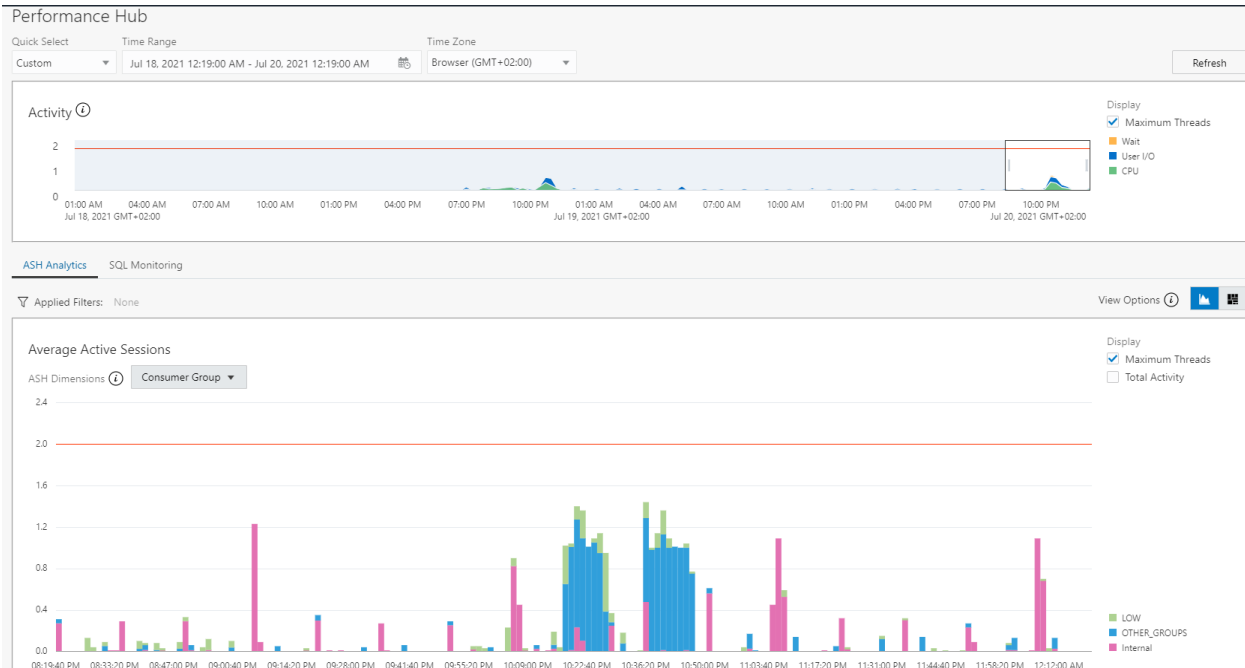


Slika 23. Charts alatom dobiven je grafički prikaz rezultata upita

Pomoću *Performance Hub* alata možete pregledavati podatke o izvedbi u stvarnom vremenu i povijesti. U *Database Management* servisu je dostupan *Performance Hub* je sa stranica s detaljima baze podataka, tako da korisnici poput administratora baze podataka mogu nadzirati vanjske baze podataka. Prozor ovog alata sastoji se od grafičkog prikaza vremenskog raspona koji se koristi za odabir vremenskog perioda svih podataka koji će se prikazati. Prva karakteristika ovog alata je birač vremenskog raspona koji je prikazan na vrhu stranice *Performance Hub*-a. Sastoji se od grafički prikazanog vremenskog polja kako je prikazano na gornjem djelu *Slike 24*. Na kartici *ASH Analytics* prikazana je aktivna povijest sesija tj. prikazani su grafikoni ASH analitike pomoću kojih se mogu istraživati ASH podaci. Ovu kartica se koristi za detaljnu analizu performansi baze podataka u više dimenzija, poput grupe potrošača, klase čekanja, SQL ID-a i korisničkog imena.



## O Oracle autonomnom skladištu podataka



Slika 24. Prikaz performansi za skladište podataka mtrstenjak\_DW\_oiri

Na Slici 24. prikazana je aktivnost baze mtrstenjak\_DW\_oiri kroz 19. Srpnja 2021. Za vrijeme korištenja *Graph* i *SQL* alata vidi se veća aktivnost označena plavom bojom, a koja prikazuje vanjske grupe koje su koristile bazu. Kasnije su interno proučavane performanse kroz *Performance Hub* i vršeni uvidi kroz *Data Insights* alate, te je ta aktivnost označena rozom bojom.

## 9. Analiza Oracle ADW sustava

Kao i prednost svih dobavljača usluga u oblaku, najveća prednost je smanjenje troškova infrastrukture, jer čak i kad se uloži u najbolju opremu, zbog ljudske pogreške uvijek može nešto poći po zlu. S pohranom u oblaku, podaci se distribuiraju među podatkovnim centrima. Tehnologija sinkronizacije omogućuje brzo povezivanje i ažuriranje podataka, uz to, kad su svi podaci pohranjeni u oblaku, točno se zna gdje se svaki podatak nalazi u bilo kojem trenutku. Gubitak podataka može značiti katastrofu za tvrtku bilo koje veličine. Povrede podataka su jako skupe, a pohrana u oblaku mnogo je sigurnija od upravljanja podatkovnim centrom na licu mjesta. Jedna od prednosti pohrane podataka u oblaku je ta što ne postoji jedna jedina točka greške, tj. podaci se sigurnosno kopiraju na nekoliko poslužitelja, pa ako jedan od njih ne uspije, podaci ostaju sigurni na drugima. Skalabilnost omogućuje plaćanje samo za količinu prostora skladištenja koja je korištena. Ako postoji potreba, povećanje oblaka dolazi s predvidljivim troškovima, uklanjajući rizik povezan s ulaganjem u dodatnu infrastrukturu za pohranu. S većom sposobnošću povećanja ili smanjenja skladišnog kapaciteta prema potrebi, organizacija postaje agilnija i konkurentnija bez obzira na industriju. Budući da infrastrukturu oblaka posjeduje i njome upravlja davatelj usluga, korisnici se mogu zabrinuti zbog nedostatka kontrole nad uslugom. Tu može pomoći licencni ugovor s krajnjim korisnikom davatelja usluga. Legitimni davatelji usluga računarstva u oblaku dopuštaju organizaciji kontrolu nad njezinim aplikacijama i podacima, ali to nje mijenja restrikciju nad mogućnosti upravljanja infrastrukturom.

Oracle autonomno skladište podataka potpuno je upravljana baza podataka koja je podešena i optimizirana za radna opterećenja skladišta podataka. Ova usluga kombinira vodeće performanse Oraclea, sama se upravlja, sama se osigurava i sama popravljiva. Kroz ovaj rad prikazane su neke od značajki, a u ovome poglavlju bit će raspravljano generalno iskustvo korištenja servisa i neke od prednosti i nedostataka.

Nakon iskustva i poznavanja AWS servisa, korišten je Oracle servis za potrebe ovoga rada. Korištenje je poprilično lako, implementirano je intuitivno grafičko korisničko sučelje, uz to, Oracle nudi bogatu dokumentaciju i pregršt tutoriala. Vrlo se brzo postavlja, za stvaranje nove baze podataka potrebno je nekoliko minuta. Podržana je vrlo bogata biblioteka statističkih i zbirnih funkcija. Sveukupno iskustvo korištenja Oracle autonomnog skladišta podataka jako je dobro, a dohvaćanje podataka je jednostavno. Doduše, zamjerka koja je primijećena je jedan tablični prostor pridružen svim korisnicima koji su stvoreni u bazi podataka.

Oracle nudi izvrsne performanse, jednostavnost korištenja, veliku skalabilnost i izvrsnu integraciju s alatima za prikaz podataka u stvarnom vremenu a time i izvještavanje u stvarnom vremenu s više korisnika i iz više izvora podataka. Jako je važna značajka sigurnosti, podaci su uvijek zaštićeni te nema potrebe za implementacijom VPN-a. S druge strane, Oracle ADW je vjerojatno manje prikladan za vrlo velike tvrtke s ogromnom količinom podataka, gdje bi došlo do kašnjenja mreže u premještanju podataka kroz samu mrežu ili za tvrtke koje već imaju jako veliko skladište podataka koje žele premjestiti u oblak, budući da autonomna baza podataka ima neka ograničenja. Postoji ograničenje u broju zapisa, te se ne može vratiti više od milijun zapisa. Ono što bi se moglo poboljšati je izvezivanje veće skupine podataka u na primjer Excel, što za sada nije izvedivo. Učitavanje podataka u oblaku podložno je brzini mreže, pa će ogromnim količinama podataka biti potrebno puno vremena da se učitaju u bazu podataka. Za ovaj rad nije bilo potrebe korištenja integracije s nekim aplikacijama ili korištenja usluge podrške, no proučavajući recenzije i povratne

## O Oracle autonomnom skladištu podataka

informacije korisnika generalno postoji prostor za poboljšanje razine integracije za povezivanje s različitim aplikacijama, te za poboljšanje usluge podrške koja se smatra sporom.

Odličan alat za edukacijske svrhe i upoznavanje budućih stručnjaka s dobavljačima usluge u oblaku. Jedina potencijalna negativna strana je cijena, no Oracle ADW nije značajno skup, troškovi se temelje na namjenskim resursima poput RAM a i CPU-a koji se mogu isključiti. Dodatno se mogu podesiti restrikcije na pojedine korisnike putem permisija u rolama. Resursi se mogu povećati ili smanjiti tijekom izvođenja ako za to ima potrebe.

Neke od značajnih prednosti i nedostataka prikazane su u tablici u nastavku.

Tablica 9. Prikaz nekih od prednosti i nedostataka Oracle autonomnog skladišta podataka

Prednosti	Nedostaci
Autonomnost	Troškovi
Sigurnost	Spora podrška
Visoka skalabilnost	Težak import iz drugih baza podataka
Pouzdanost	Kompleksnost inicijalne konfiguracije
Intuitivnost	Spor eksport i import velikih objekata
Odlične performanse	Ograničenje od 30 znakova za ime objekta
Stabilnost	Nemogućnost kontrole veličine memorije instanci

Autonomnost, tj. samostalnost najjača je prednost Oracleovog skladišta podataka jer je jedino cjelovito rješenje koje koristi konvergiranu bazu podataka koja pruža ugrađenu podršku za multimodelne podatke i višestruka radna opterećenja. Uključuje ugrađene alate za samoposluživanje za poboljšanje produktivnosti analitičara, podatkovnih znanstvenika i programera. Velika prednost je i *Oracle Data Safe* sigurnosni kontrolni centar zasnovan na oblaku za Oracle skladišta podataka i to besplatno. Ovaj servis pruža bitne mogućnosti zaštite podataka, pomaže razumjeti osjetljivost podataka, procijeniti rizike po podatke, prikriti osjetljive podatke, implementirati i nadzirati sigurnosne kontrole, procijeniti sigurnost korisnika i pratiti aktivnosti korisnika. Autonomno skladište podataka nudi i automatsko skaliranje čime se može koristiti puno više CPU resursa. Ako radno opterećenje zahtijeva dodatne resurse, baza podataka automatski dodaje resurse bez ikakve ručne intervencije što još više smanjuje potrebu za upravljanjem i praćenjem. Uz navedeno, korisničko grafičko sučelje je jednostavno i jasno za korištenje, iako za potrebe ovoga rada nisu korištene komplicirane konfiguracije koje se navode kao jedan od nedostataka ovog servisa.

Jedan od nedostataka autonomnog skladišta podataka su troškovi, plaća se hodu i to samo usluge koje se koriste. Ako postoji potreba za dodatnim sredstvima, mogu se povećati i plaća se samo za njih. Nema potrebe brinuti se oko plaćanja nečega što se ne koristi, ali za velike količine podataka i puno resursa troškovi budu značajno visoki. Jedan od većih nedostataka je i spora korisnička podrška, te kompleksnost inicijalne konfiguracije sustava. Također, u realnom poslovnom svijetu ograničenje od 30 znakova u imenovanju je nedostatak, pogotovo kada se dinamički generiraju imena na temelju vremenske oznake uz imenovanje funkcionalnosti i slično.

Bez obzira koji, dobavljač usluge u oblaku trebao bi biti sadržan u izvedbenom plan una fakultetima tehničkih usmjerenja jer potrebe tržišta se značajno razvijaju u tom smjeru.

## 10. Zaključak

Do prije nekoliko godina Amazon AWS bio je jedini održivi pružatelj usluga u oblaku hiperrazmjera, Azure je bio u usponu, Google je tek počinjao, dok Oracle nije bio na toj razini. Svi su navedeni dobavljači postigli izuzetan napredak do te mjere da korisnici imaju mnogo održivih alternativa, što je pozitivno, jer konkurencija potiče inovacije i niže troškove. Oracle je daleko napredovao od Cloud V1.0 i bolji je proizvod od percepcije tržišta trenutno, ali njegova popularnost je u usponu. Oracle sada kombinira fleksibilnost oblaka sa snagom strojnog učenja za upravljanje podacima, a izgrađen je na temelju tehničkih inovacija koje je Oracle razvio tijekom više od tri desetljeća zadovoljavajući potrebe tisuća poslovnih korisnika širom svijeta.

Većina dobavljača danas nudi skladište podataka u oblaku, a Oracleova nova generacija autonomnih skladišta nudi korisnicima jedinstvenu, konvergiranu bazu podataka u oblaku koja bi trebala zadovoljiti većinu poslovnih potreba, eliminirajući potrebu za kupnjom specijaliziranih baza podataka i potrebu za integracijom podataka iz više izvora. Oracle je skladište podataka u oblaku učinio samostalnim, što je uklonilo puno operativnih administrativnih troškova i ručnog rada, te dovelo do dugoročne uštede novca, pružajući bolju pouzdanost i bolju razinu usluge.

Oracle ADW usluga je skladišta podataka u oblaku. Automatiziranjem opskrbe, konfiguriranja, skaliranja, sigurnosnog kopiranja i održavanja skladišta podataka, pruža jednostavnu i uvijek dostupnu uslugu visokih performansi. Uključuje ugrađenu analitiku, poput strojnog učenja u bazi podataka, te sveobuhvatne mogućnosti sigurnosti i privatnosti za zaštitu podataka korisnika. Neke od najznačajnijih prednosti Oracle autonomnog skladišta podataka su jednostavnost korištenja, intuitivnost, jednostavna konfiguracija, dobre performanse i visoka skalabilnost, samostalno održavanje i automatsko održavanje, brze ETL značajke omogućuju brzo učitavanje podataka, nema brige o sigurnosnim kopijama ili održavanju jer je u oblaku. Negativna karakteristika ove usluge je potencijalno visoka cijena ako se koriste sve značajke.

Konkurentnim davatelji usluga oblaka nedostaje većina naprednih značajki Oracleovih ponuda, zapravo nijedan pružatelj usluge u oblaku ne nudi autonomne mogućnosti koje trenutno nudi Oracle. Što je još važnije, ne postoji poznato potpuno automatizirano, potpuno upravljano skladište podataka poput Oracleovog. Najnovije izdanje Oracle ADW uključuje mnoge inovacije, poput širokog skup mogućnosti koje analitičarima, građanskim znanstvenicima i razvojnim programerima olakšavaju iskorištavanje prvog i jedinog samostalnog skladišta podataka u oblaku u industriji, ali i značajke koje pružaju dublju analitiku i čvršću integraciju jezera podataka. Unatoč tome, čini se vjerojatnim da će drugi pružatelji usluga u oblaku razviti konkurentnija rješenja, pa Oracle mora i dalje ostati inovativan.

Skladišta podataka obično sadrže velike količine podataka koje se skupno obrađuju ili prenose u bazu podataka, te se često oslanjaju na sažeti prikaz podataka i paralelni SQL kako bi bio omogućen brz odgovor, a Oracle ADW posebno je prilagođen ovim slučajevima upotrebe. Odličan je izbor za tvrtke jer osigurava sigurnost podataka, poprilično je automatiziran alat, čime se umanjuje potreba za ljudskim intervencijama, te se smanjuje mogućnost ljudske pogreške. Intuitivan je alat uz dobru dokumentaciju i pregršt video materijala dostupnih na webu pa je time pogodan i u edukacijske svrhe, te nudi probnu verziju za korisnike koji žele isprobati kako cijeli servis funkcionira.

## Popis literature

- [1] Amazon Redshift. Preuzeto 14. srpnja 2021. <https://aws.amazon.com/redshift/?whats-new-cards.sort-by=item.additionalFields.postDateTime&whats-new-cards.sort-order=desc>.
- [2] Autonomous Data Warehouse: New innovations for data analysts, citizen data scientists and LOB developers. Oracle Database Insider. Izvor prema stanju od 17. ožujka 2021. Preuzeto 20. srpnja 2021. <https://blogs.oracle.com/database/post/autonomous-data-warehouse-new-innovations-for-data-analysts-citizen-data-scientists-and-lob-developers>.
- [3] Everything You Need to Know About Cloud Computing. Smart Growth. Izvor prema stanju od 10. lipnja 2021. Preuzeto 17. srpnja 2021. <https://smartgrowthvermont.org/everything-you-need-to-know-about-cloud-computing/>.
- [4] Few interesting facts about Oracle ADB, Redshift and Snowflake. Dontcheff. Izvor prema stanju od 14. siječnja 2019. Preuzeto 19. srpnja 2021. <https://juliandontcheff.wordpress.com/2019/01/14/few-interesting-facts-about-oracle-adb-redshift-and-snowflake/>.
- [5] Oracle Autonomous Data Warehouse. Izvor prema stanju od 2021. Preuzeto 12. srpnja 2021. <https://www.oracle.com/autonomous-database/autonomous-data-warehouse/>.
- [6] Oracle Autonomous Database Strategy. Izvor prema stanju od lipnja 2020. Preuzeto 14. srpnja 2021. <https://www.oracle.com/a/ocom/docs/database/oracle-autonomous-database-strategy-wp.pdf>.
- [7] Oracle Autonomous Data Warehouse: The Future of Data Warehousing in the Cloud. Preuzeto 14. srpnja 2021. <https://explore.oracle.com/adw-ocom/future-of-data>.
- [8] Oracle Autonomous Data Warehouse vs Amazon Redshift. Preuzeto 19. srpnja 2021. <https://www.trustradius.com/compare-products/oracle-autonomous-data-warehouse-vs-redshift>.
- [9] Putting Data Security and Protection First. Izvor prema stanju od 2020. Preuzeto 14. srpnja 2021. <https://www.oracle.com/webfolder/s/assets/ebook/security-first/index.html>.

O Oracle autonomnom skladištu podataka

[10] Računarstvo u oblaku. Izvor prema stanju od 21. lipnja 2021. Preuzeto 16. srpnja 2021.  
[https://hr.wikipedia.org/wiki/Ra%C4%8Dunarstvo\\_u\\_oblaku](https://hr.wikipedia.org/wiki/Ra%C4%8Dunarstvo_u_oblaku).

[11] Suhag, Amit. “Oracle Autonomous Database: All You Need to Know”. Cloud Management Insider. Izvor prema stanju od 25. svibnja 2020. Preuzeto 13. srpnja 2021.  
<https://www.cloudmanagementinsider.com/oracle-autonomous-database-all-you-need-to-know/>.

[12] Tickit podatkovni skup. Preuzeto 17. srpnja 2021.  
<https://docs.aws.amazon.com/redshift/latest/gsg/samples/tickitdb.zip>.

[13] Tickit diagram. Dataedo. Preuzeto 18. srpnja 2021.  
[https://dataedo.com/samples/html/Tickit/doc/Tickit\\_16/modules/Tickit\\_diagram\\_136/module.html](https://dataedo.com/samples/html/Tickit/doc/Tickit_16/modules/Tickit_diagram_136/module.html).

[14] Tippe, Sagar. “Oracle Business Intelligence 11g”. Izvor prema stanju od 08. ožujka 2029. Preuzeto 13. srpnja 2021. <https://obieenow.blogspot.com/2019/03/introduction-to-autonomous-data-warehouse.html>.

[15] Wheeler, Patrick. “Introducing Autonomous Data Tools”. Oracle Database Insider. Izvor prema stanju od 17. ožujka 2021. Preuzeto 17. srpnja 2021.  
<https://blogs.oracle.com/database/post/introducing-autonomous-data-tools>.

[16] Zebari, Rizgar. “Cloud computing services (IaaS, PaaS, and SaaS)”. Preuzeto 15. srpnja 2021. [https://www.researchgate.net/figure/Cloud-computing-services-IaaS-PaaS-and-SaaS\\_fig1\\_342492294](https://www.researchgate.net/figure/Cloud-computing-services-IaaS-PaaS-and-SaaS_fig1_342492294).

## Popis slika

- Slika 1. Koncept računarstva u oblaku
- Slika 2. IaaS, PaaS i SaaS
- Slika 3. Oracle skladište podataka logo
- Slika 4. Tri glavna tipa korisnika Oracle skladišta podataka
- Slika 5. Podatkovni alati dostupni kroz Oracle DB servis
- Slika 6. Glavne razlike između Oracl-a i Amazon AWS-a
- Slika 7. Pokretanje besplatnog roka korištenja Oracle servisa
- Slika 8. Prva stranica nakon kreiranja računa na Oracle servisu je *Dashboard* tj. nadzorna ploča
- Slika 9. Popis svih dostupnih Oracle servisa
- Slika 10. Kreiranje skladišta podataka kroz servis baza podataka (*Autonomous Database*)
- Slika 11. Kreirano skladište podataka naziva *mtrstenjak\_DW\_oiri* u dostupnom stanju
- Slika 12. Dostupni podatkovni alati
- Slika 13. Prozor za postavljanje parametara i tipova podataka prije njihova učitavanja u katalog
- Slika 14. Prikaz kataloga sa nekim od kreiranih tablica
- Slika 15. Dio *Event* tablice
- Slika 16. Prikaz statistike za *Event* tablicu
- Slika 17. Prikaz DDL-a *Event* tablice
- Slika 18. Kreiran model naziva *EVENTS\_MODEL\_AV*
- Slika 19. Kreirane mjere na temelju kojih će poslovni model biti proučavan
- Slika 20. Grafički prikaz broja održanih događaja po mjesu uz označena značajna odstupanja
- Slika 21. Izgled sučelja unutar *SQL Page* alata
- Slika 22. Upit nad *Tickit* skupom podataka sa pripadajućim izlaznim rezultatom
- Slika 23. *Charts* alatom dobiven je grafički prikaz rezultata upita
- Slika 24. Prikaz performansi za skladište podataka *mtrstenjak\_DW\_oiri*

## Popis tablica

Tablica 1. Uporedba Oracle ADW i AWS Redshifta prema odabranim značajkama

Tablica 2. Prikaz detalja o tablici *Category*

Tablica 3. Prikaz detalja o tablici *Date*

Tablica 4. Prikaz detalja o tablici *Event*

Tablica 5. Prikaz detalja o tablici *Venue*

Tablica 6. Prikaz detalja o tablici *Users*

Tablica 7. Prikaz detalja o tablici *Listing*

Tablica 8. Prikaz detalja o tablici *Sales*

Tablica 9. Prikaz nekih od prednosti i nedostataka Oracle autonomnog skladišta podataka

## Popis grafova

Graf 1. Veze između tablica Ticket skupa podataka