ALATI ZA OBRADU PODATAKA

Profesionalne uloge u svetu podataka

Ko projektuje, ko priprema, ko analizira i ko uvodi modele u realan sistem?

- DA **projektuje**
- DE **priprema**
- BI **objašnjava**
- ML **primenjuje**

Predavač: dr Dušan Stefanović

Isti naziv posla može značiti različite zadatke

ZAŠTO NASTAJE ZABUNA

DATA ANALYST – FIRMA A

SQL upiti
Excel analiza
Power BI izveštaji

fokus: izveštavanje

DATA ANALYST – FIRMA B

Python i statistika
segmentacija kupaca
jednostavne predikcije

fokus: naprednija analiza

Naslov oglasa nije dovoljan za procenu sadržaja posla.

Oglas za posao čitamo kroz četiri pitanja

KAKO PREPOZNATI ULOGU

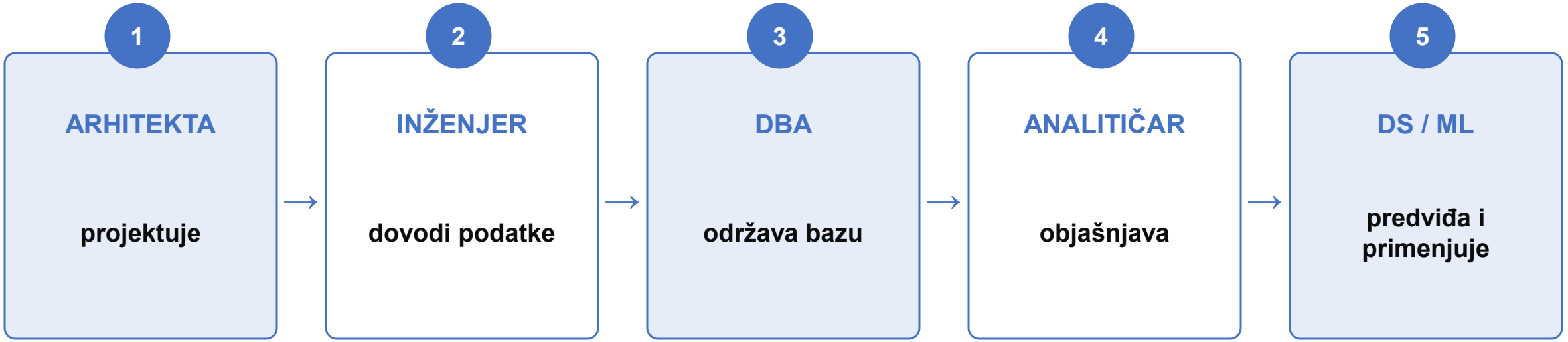
1	2	3	4
ZADATAK	**ALAT**	**REZULTAT**	**ODGOVORNOST**
Šta osoba svakodnevno radi?	SQL, Python, Power BI ili cloud?	Tabela, dashboard, model ili servis?	Razvoj, analiza ili održavanje?

Primer: „održava pipeline i ETL skripte” ukazuje na data engineering, čak i ako naslov oglasa glasi Data Analyst.

Opis odgovornosti ima veću vrednost od zvučnog naziva pozicije.

Tim pretvara sirove podatke u odluku

ZAJEDNIČKI TOK RADA



Vrednost nastaje tek kada se rezultati uloga međusobno povežu.

Data architect projektuje organizaciju podataka

ARHITEKTURA PODATAKA

DATA ARCHITECT

arhitekta podataka

GLAVNO PITANJE

Kako podatke organizovati za celu organizaciju?

TIPIČNI ZADACI

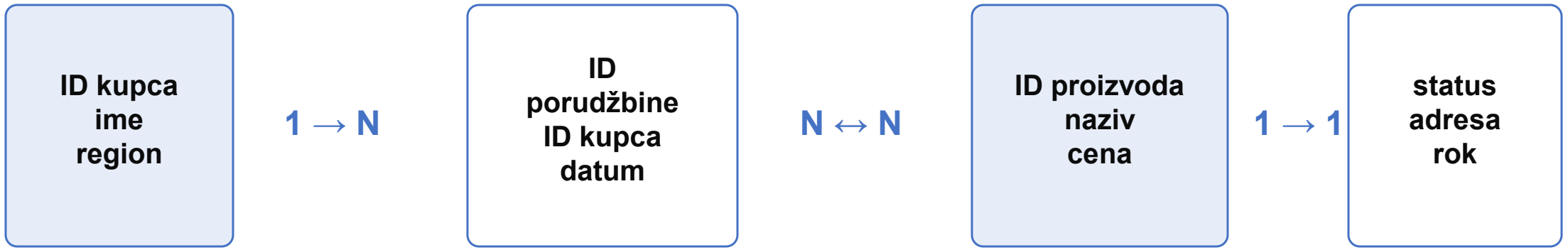
1. definiše modele i veze
2. planira integracije i standarde
3. usklađuje bezbednost i potrebe korisnika

Rezultat rada: **model podataka, standardi i arhitekturni plan**

Arhitekta donosi odluke koje usmeravaju kasniju izgradnju sistema.

Primer: arhitektura internet prodavnice

MODEL PRE IMPLEMENTACIJE



Arhitekta određuje identifikatore, veze, pravila i vlasništvo nad podacima pre izrade tokova.

Dobar model sprečava nedoslednosti i olakšava kasniju analizu.

Data engineer gradi pouzdane tokove podataka

INŽENJERSTVO PODATAKA

DATA ENGINEER

inženjer podataka

GLAVNO PITANJE

Kako podatke pouzdano dovesti i pripremiti?

TIPIČNI ZADACI

1. povezuje izvore i odredišta
2. razvija ETL/ELT tokove
3. automatizuje kontrole kvaliteta

Rezultat rada: **osvežen, organizovan i dostupan skup podataka**

Analitičar može verovati podacima samo ako je tok pouzdan.

ETL pretvara više izvora u analitički skup

PRIMER TOKA PODATAKA



Rezultat data engineeringa nije samo tabela, već ponovljiv proces.

Big Data uloge rade u većoj razmeri

DISTRIBUIRANI SISTEMI

BIG DATA ARCHITECT

projektuje distribuirano čuvanje
obradu u realnom vremenu
skaliranje i otpornost

odluke o platformi

BIG DATA ENGINEER

gradi distribuirane pipeline-ove
obrađuje velike tokove
prati izvršavanje poslova

implementacija i rad tokova

Primer: telekom obrađuje milione događaja uređaja koristeći cloud, Kafka, Spark ili srodne tehnologije.

„Big Data" opisuje obim i arhitekturu, a ne potpuno odvojenu profesiju.

DBA čuva bazu dostupnom, bezbednom i brzom

ADMINISTRACIJA BAZE

DBA

administrator baze podataka

GLAVNO PITANJE

Kako obezbediti pouzdan rad baze?

TIPIČNI ZADACI

1. upravlja pristupom i zaštitom
2. pravi backup i plan oporavka
3. prati i optimizuje performanse

Rezultat rada: **stabilna, zaštićena i dostupna baza**

DBA održava sistem; arhitekta projektuje širu organizaciju podataka.

BI analyst objašnjava šta se dogodilo

POSLOVNA INTELIGENCIJA

BI ANALYST

BI analitičar

GLAVNO PITANJE

Šta se dogodilo i gde je nastao problem?

TIPIČNI ZADACI

1. definiše i prati KPI pokazatelje
2. analizira istorijske podatke
3. komunicira zaključke menadžmentu

Rezultat rada: **izveštaj, dashboard i poslovni uvid**

BI analitičar pretvara istorijske podatke u razumljivu poslovnu priču.

BI consultant prenosi znanje između projekata

KONSULTANTSKA ULOGA

BI CONSULTANT

BI konsultant

GLAVNO PITANJE

Kako organizacija može uvesti ili unaprediti BI?

TIPIČNI ZADACI

1. analizira zahteve i postojeće stanje
2. predlaže platformu i pokazatelje
3. implementira rešenje i obučava korisnike

Rezultat rada: **predlog, početno BI rešenje i prenos znanja**

Konsultant je često spoljni BI analitičar ili developer sa širom projektnom ulogom.

BI developer tehnički gradi sistem izveštavanja

RAZVOJ BI REŠENJA

BI DEVELOPER

programer BI rešenja

GLAVNO PITANJE

Kako zahtev pretvoriti u pouzdan BI sistem?

TIPIČNI ZADACI

1. piše SQL i priprema modele podataka
2. definiše poslovne mere i pravila
3. razvija dashboard i osvežavanje

Rezultat rada: **BI model, kontrolna tabla i automatizovano osvežavanje**

BI developer spaja bazu, poslovna pravila i vizuelni prikaz.

Data analyst odgovara na konkretna pitanja podacima

ANALIZA PODATAKA

DATA ANALYST

analitičar podataka

GLAVNO PITANJE

Šta podaci govore o konkretnom pitanju?

TIPIČNI ZADACI

1. prikuplja, čisti i istražuje podatke
2. primenjuje analizu i vizualizaciju
3. prikazuje zaključak različitim korisnicima

Rezultat rada: **analiza, vizuelni prikaz i preporuka**

Analitičar povezuje tehnički rezultat sa pitanjem koje organizacija želi da reši.

BI analyst i data analyst imaju različito težište

POREĐENJE ANALITIČKIH ULOGA

Pitanje	BI analyst	Data analyst
Glavni fokus	poslovni pokazatelji i izveštaji	šire istraživanje konkretnog problema
Tipični podaci	standardizovani i istorijski	različiti izvori i ad hoc skupovi
Tipičan rezultat	dashboard i redovni izveštaj	analiza, vizualizacija i preporuka
Alati	SQL i BI platforma	SQL, Excel, Python/R i vizualizacija

U manjim organizacijama jedna osoba često obavlja obe uloge.

Data scientist razvija modele i proverava pretpostavke

NAUKA O PODACIMA

DATA SCIENTIST

naučnik za podatke

GLAVNO PITANJE

Šta možemo predvideti ili optimizovati?

TIPIČNI ZADACI

1. istražuje obrasce i bira promenljive
2. razvija i validira statističke/ML modele
3. prevodi rezultat u poslovnu preporuku

Rezultat rada: **proveren model, predikcije i objašnjenje rezultata**

Model vredi samo ako rešava jasno definisan problem i može da se proveri.

ML engineer pretvara model u pouzdan servis

MAŠINSKO UČENJE U PRODUKCIJI

ML ENGINEER

inženjer mašinskog učenja

GLAVNO PITANJE

Kako model bezbedno koristiti u realnom sistemu?

TIPIČNI ZADACI

1. postavlja i povezuje model sa aplikacijom
2. obezbeđuje brzinu, skaliranje i testiranje
3. prati kvalitet i organizuje ponovno obučavanje

Rezultat rada: **produkcioni ML servis koji se može nadgledati**

Uspešan eksperiment postaje koristan tek kada pouzdano radi nad novim podacima.

Model povezuje data scientista i ML engineera

OD EKSPERIMENTA DO PRODUKCIJE

Pitanje	Data scientist	ML engineer
Glavni cilj	napraviti i proveriti koristan model	obezbediti pouzdan rad modela
Okruženje	eksperiment i analiza	produkcioni sistem
Mere uspeha	tačnost, greška, poslovni efekat	brzina, dostupnost, stabilnost i kvalitet
Tipičan rezultat	model i evaluacija	servis, monitoring i proces ažuriranja

Najbolji rezultat nastaje kada se model razvija uz svest o produkcionim uslovima.

Use case: kako zadržati kupce

ZAJEDNIČKI POSLOVNI PRIMER

PROBLEM

Kupci prestaju da kupuju, ali kompanija ne zna ko je rizičan niti kada treba reagovati.

1. **povezati podatke**
2. **otkriti obrazac**
3. **predvideti rizik**
4. **pokrenuti akciju**
5. **izmeriti efekat**

Jedan poslovni problem zahteva više različitih profesionalnih uloga.

Svaka uloga preuzima jedan deo churn problema

PODELA ODGOVORNOSTI

1	Data architect	povezuje kupca, kupovine, sajt i reklamacije
2	Data engineer	gradi svakodnevni tok i proverava kvalitet
3	DBA	štiti bazu i održava njenu dostupnost
4	BI/Data analyst	otkriva grupe sa najvećim odlaskom
5	Data scientist	razvija model verovatnoće odlaska
6	ML engineer	postavlja model i prati njegovo ponašanje

Podela rada sprečava da važan deo sistema ostane bez vlasnika.

Rezultat jedne uloge postaje ulaz sledeće

PREDAJA ODGOVORNOSTI



Svaki izlaz mora imati poznatu strukturu, vlasnika, vreme osvežavanja i proveru kvaliteta.

Nejasna predaja između uloga stvara greške čak i kada je svaki alat dobar.

Interesovanja usmeravaju izbor uloge

ORIJENTACIJA KARIJERE

Ako vas najviše zanima…	Moguća uloga	Početni fokus
struktura i povezivanje sistema	Data architect	modelovanje podataka i arhitektura
automatizacija i pipeline-ovi	Data engineer	SQL, Python i ETL
stabilnost i bezbednost baze	DBA	SQL sistem, backup i performanse
izveštaji i poslovna komunikacija	BI/Data analyst	SQL, statistika i vizualizacija
modeli i eksperimentisanje	Data scientist	Python/R, statistika i ML
softver i rad modela u produkciji	ML engineer	Python, API, cloud i MLOps

Početna uloga nije trajna granica; veštine se međusobno dopunjuju.

Sedam poruka razjašnjava profesionalne uloge

PREGLED LEKCIJE

1. Arhitekta projektuje širu organizaciju podataka.
2. Data engineer gradi ponovljive tokove.
3. DBA održava bazu bezbednom i dostupnom.
4. BI tim pretvara istorijske podatke u uvide.
5. Data analyst odgovara na konkretna pitanja.
6. Data scientist razvija i proverava modele.
7. **ML engineer uvodi i održava model u produkciji.**

Prilikom čitanja oglasa posmatrajte zadatak, alat, rezultat i odgovornost.

Pitanje 1

PROVERA ZNANJA

Koji rezultat je najtipičniji za data architecta?

- **A** Produkcioni ML servis
- **B** Model podataka i arhitekturni plan
- **C** Dnevni prodajni dashboard
- **D** Rezervna kopija baze

Izaberite jedan odgovor i obrazložite zašto ostali nisu odgovarajući.

Pitanje 2

PROVERA ZNANJA

Koja aktivnost najjasnije pripada data engineeru?

A Razvoj ETL toka koji se izvršava svake noći

B Izbor boje grafikona za upravu

C Odobravanje marketinškog budžeta

D Pisanje zaključka ankete

Izaberite jedan odgovor i obrazložite zašto ostali nisu odgovarajući.

Pitanje 3

PROVERA ZNANJA

Ko je prvenstveno odgovoran za backup i performanse baze?

A BI consultant

B Data scientist

C Database administrator

D ML engineer

Izaberite jedan odgovor i obrazložite zašto ostali nisu odgovarajući.

Pitanje 4

PROVERA ZNANJA

Šta najbolje razlikuje ML engineera od data scientista?

A ML engineer ne koristi modele

B Data scientist održava servere baze

C ML engineer uvodi, skalira i prati model u produkciji

D Data scientist pravi samo grafikone

Izaberite jedan odgovor i obrazložite zašto ostali nisu odgovarajući.

Pitanje 5

PROVERA ZNANJA

Kako najpouzdanije procenjujemo sadržaj oglasa za posao?

A Samo prema naslovu pozicije

B Prema veličini logotipa firme

C Prema zadacima, alatima, rezultatu i odgovornosti

D Prema broju stranih reči

Izaberite jedan odgovor i obrazložite zašto ostali nisu odgovarajući.

Praktičan zadatak: upravljanje zalihama

SAMOSTALNA PRIMENA

SCENARIO

Lanac prodavnica želi da predvidi proizvode kojima preti nestanak i automatski predloži novu porudžbinu.

1. **Ko projektuje model proizvoda, prodavnica i dobavljača?**
2. **Ko prikuplja prodaju iz svih prodavnica?**
3. **Ko održava bazu i rezervne kopije?**
4. **Ko analizira uzroke prethodnih nestašica?**
5. **Ko razvija, a ko postavlja model u sistem?**

Dobar odgovor navodi ulogu, njen rezultat i kome se taj rezultat predaje.