

Predgovor .....	xvii
-----------------	------

---

## Deo I Temelji i gradivni blokovi

<b>1. Opis inženjerstva podataka .....</b>	<b>3</b>
Šta je inženjerstvo podataka?	3
Definicija inženjerstva podataka	4
Životni ciklus inženjerstva podataka	5
Evolucija inženjera podataka	6
Inženjerstvo podataka i nauka o podacima	11
Veštine i aktivnosti inženjerstva podataka	13
Zrelost podataka i inženjer podataka	14
Pozadina i veštine inženjera podataka	17
Poslovne odgovornosti	18
Tehničke odgovornosti	19
Kontinuum uloga inženjera podataka, od A do B	22
Inženjeri podataka unutar organizacije	23
Inženjeri podataka okrenuti prema unutra u odnosu na one okrenute prema spolja	23
Inženjeri za obradu podataka i ostale tehničke uloge	25
Inženjeri podataka i poslovno liderstvo	29
Zaključak	32
Dodatni izvori	33
<b>2. Životni ciklus inženjerstva podataka .....</b>	<b>35</b>
Šta je životni ciklus inženjerstva podataka?	35
Životni ciklus podataka naspram ciklusa inženjerstva podataka	36
Generisanje: izvorni sistemi	37
Skladištenje	40
Unos	42
Transformacija	45
Serviranje podataka	47

Glavni podtokovi kroz ciklus inženjerstva podataka	52
Bezbednost	52
Menadžment podataka	53
DataOps	63
Arhitektura podataka	67
Orkestracija	68
Softversko inženjerstvo	69
Zaključak	72
Dodatni izvori	73
<b>3. Projektovanje dobre arhitekture podataka.....</b>	<b>75</b>
Šta je arhitektura podataka?	75
Definisanje arhitekture preduzeća	76
Definicija arhitekture podataka	79
„Dobra“ arhitektura podataka	80
Principi dobre arhitekture podataka	81
Princip 1: Mudro birajte zajedničke komponente	82
Princip 2: Planirajte za slučaj neuspeha	83
Princip 3: Projektujte za skalabilnost	83
Princip 4: Arhitektura je vođstvo	84
Princip 5: Uvek budite usmereni na arhitekturu	85
Princip 6: Gradite sisteme sa labavom spregom	85
Princip 7: Donosite reverzibilne odluke	87
Princip 8: Neka vam bezbednost bude prioritet	87
Princip 9: Prihvatite FinOps	89
Glavni koncepti arhitekture	91
Domeni i usluge	91
Distribuirani sistemi, skalabilnost i dizajniranje za slučaj neuspeha	92
Čvrsta naspram labave povezanosti: nivoi, monoliti i mikroservisi	94
Pristup korisnika: pojedinačni naspram višekorisničkog	99
Arhitektura zasnovana na događajima	99
Braunfeld nasuprot greenfield projekata	100
Primeri i tipovi arhitekture podataka	102
Skladište podataka	102
Jezero podataka	106
Konvergencija, jezera podataka nove generacije i platforma podataka	107
Savremen stek podataka	108
Lambda arhitektura	109
Kappa arhitektura	110
Model protoka podataka i ujedinjavanje paketa i strimovanja	110
Arhitektura za IoT	111
Mreža podataka	114
Ostali primeri arhitekture podataka	115

Ko učestvuje u dizajniranju arhitekture podataka?	116
Zaključak	116
Dodatni izvori	116
<b>4. Izbor tehnologija tokom ciklusa inženjerstva podataka .....</b>	<b>121</b>
Veličina tima i sposobnosti	122
Brzina izlaska na tržište	123
Interoperabilnost	123
Optimizacija troškova i poslovna vrednost	124
Ukupan trošak posedovanja	124
Ukupan trošak mogućnosti posedovanja	125
FinOps	126
Sadašnjost naspram budućnosti: nepromenjive naspram privremenih tehnologija	127
Naša preporuka	128
Lokacija	129
U lokalu	129
Oblak	130
Hibridni oblak	134
Multiklaud	135
Decentralizovano: blokčejn i edge računarstvo	136
Naša preporuka	136
Argumenti za repatrijaciju na oblak	137
Izgradnja naspram kupovine	139
Softver otvorenog koda	140
Vlasnički zatvoreni vrtovi	144
Naša preporuka	145
Monolit naspram modularnog	146
Monolit	146
Modularnost	147
Distribuiran monolitni obrazac	149
Naš savet	149
Bez servera naspram servera	150
Sistemi bez servera	150
Kontejneri	151
Kako proceniti server nasuprot serverless	152
Naši saveti	154
Optimizacija, performanse i ratovi benchmarka	155
Veliki podaci...za 1990-te	156
Besmisleno poređenje troškova	156
Asimetrična optimizacija	156
Caveat emptor (Neka kupac bude oprezan)	156
Podtokovi i njihov uticaj na izbor tehnologija	157
Menadžment podataka	157
DataOps	157

Arhitektura podataka	158
Primer orkestracije: Airflow	158
Inženjerstvo softvera	159
Zaključak	159
Dodatni izvori	159

---

## Deo II Životni ciklus inženjerstva podataka

<b>5. Generisanje podataka u izvornim sistemima .....</b>	<b>163</b>
Izvori podataka: kako nastaju podaci?	164
Izvorni sistemi: osnovne ideje	164
Fajlovi i nestrukturisani podaci	165
API	165
Aplikacione baze podataka (OLTP sistemi)	165
Sistem za onlajn analitičku obradu podataka	167
Praćenje promena podataka	168
Logovi	168
Logovi baze podataka	170
CRUD	170
Samo-umetanje	171
Poruke i tokovi	171
Tipovi vremena	173
Praktični detalji izvornog sistema	174
Baze podataka	174
API	183
Deljenje podataka	185
Izvori podataka treće strane	186
Redovi čekanja poruka i platforme za strimovanje događaja	186
Sa kim ćete saradivati	190
Podtokovi i njihov uticaj na izvorne sisteme	192
Bezbednost	192
Menadžment podataka	192
DataOps	193
Arhitektura podataka	194
Orkestracija	195
Softversko inženjerstvo	196
Zaključak	196
Dodatni izvori	197
<b>6. Skladištenje .....</b>	<b>199</b>
Osnovni elementi skladištenja podataka	201
Magnetni disk	201
Memorija čvrstog stanja	203

Memorija sa nasumičnim pristupom	204
Mreža i CPU	205
Serijalizacija	206
Kompresija	206
Keširanje	207
Sistemi skladištenja podataka	208
Skladištenje na jednom serveru naspram distribuiranog skladištenja	208
Labava naspram čvrste konzistentnosti	208
Skladištenje datoteka	210
Blok skladištenje	212
Skladištenje objekata	216
Skladištenje bazirano na kešu i na memoriji	222
Hadoop distribuiran fajl sistem	222
Skladištenje strimovanja	223
Indeksi, particionisanje i klasterovanje	224
Apstrakcije skladištenja u inženjerstvu podataka	226
Skladište podatka	226
Jezero podataka	227
Skladište jezera podataka	227
Platforme podataka	228
Arhitektura skladištenja od toka do paketa	229
Velike ideje i trendovi u skladištenju	229
Katalog podataka	229
Deljenje podataka	230
Šema	231
Odvojenost obrade podataka od skladištenja	231
Životni ciklus čuvanja podataka i zadržavanje podataka	235
Pojedinačno naspram višekorisničkog skladištenja	238
Sa kime ćete raditi	239
Podtokovi	240
Bezbednost	240
Menadžment podataka	240
DataOps	241
Arhitektura podataka	241
Orkestracija	242
Inženjerstvo softvera	242
Zaključak	242
Dodatni izvori	243

## **7. Unošenje podataka .....245**

Šta je unošenje podataka?	246
Ključna inženjerska razmatranja za fazu unosa	247
Ograničeni naspram neograničenih podataka	248
Frekvencija	249

Sinhron naspram asinhronog unosa podataka	250
Serijalizacija i deserijalizacija	251
Propusnost i skalabilnost	252
Pouzdanost i trajnost	252
Sadržaj paketa podataka	253
Push nasuprot Pull nasuprot Poll obrazaca	256
Paketni unos podataka	257
Snimak tekućeg stanja ili diferencijalna ekstrakcija	258
Izvoz i unošenje podataka zasnovani na datotekama	258
ETL naspram ELT	258
Unosi, ažuriranja i veličina paketa	259
Migracija podataka	259
Unos poruka i tokova podataka	260
Evolucija šeme	260
Podaci koji kasne u dolasku	260
Redosled i višestruka isporuka	261
Reprodukcija	261
TTL	261
Veličina poruke	262
Otklanjanje grešaka i redovi čekanja za neispravne poruke	262
Povlačenje i prosleđivanje poruka potrošačima	262
Lokacija	263
Načini unošenja podataka	263
Direktna veza sa bazom podataka	263
Snimanje podataka o promenama	264
API	267
Redovi čekanja poruka i platforme za emitovanje događaja	268
Upravljeni konektori podataka	269
Premeštanje podataka putem objektnog skladištenja	270
EDI	270
Baze podataka i izvoz fajlova	270
Praktični problemi sa uobičajenim formatima fajlova	271
Shell	271
SSH	272
SFTP i SCP	272
Veb udice	272
Veb interfejs	273
Veb skreping	273
Uređaji za transfer podataka	274
Deljenje podataka	275
Sa kim ćete saradivati	275
Uzvodne zainteresovane strane	275
Zainteresovane strane na nižem nivou	276

Podtokovi	277
Bezbednost	277
Upravljanje podacima	277
DataOps	279
Orkestracija	281
Inženjerstvo softvera	282
Zaključak	282
Dodatni izvori	282
<b>8. Upiti, modelovanje i transformacija .....</b>	<b>283</b>
Upiti	284
Šta je upit?	284
Život upita	286
Optimizator upita	287
Poboljšanje performansi upita	287
Upiti nad tokovima podataka	293
Modelovanje podataka	300
Šta je model podataka?	300
Konceptualni, logički i fizički modeli podataka	301
Normalizacija	303
Tehnike za modelovanje paketnih analitičkih podataka	307
Transformacije	322
Paketne transformacije	323
Materijalizovani prikazi, federacija i virtualizacija upita	337
Strimovanje transformacija i obrada	340
S kim ćete raditi	343
Nadređene zainteresovane strane	343
Zainteresovane strane nizvodno	344
Podstrujanja	344
Bezbednost	344
Menadžment podataka	345
DataOps	346
Arhitektura podataka	346
Orkestracija	347
Inženjerstvo softvera	347
Zaključak	348
Dodatni izvori	349
<b>9. Serviranje podataka za analitiku, mašinsko učenje i reverzan ETL .....</b>	<b>351</b>
Opšta razmatranja u vezi serviranja podataka	352
Poverenje	352
Šta je slučaj upotrebe i ko je korisnik?	354
Data proizvodi	354

Samoposluživanje ili ne?	355
Definicije podataka i logika	356
Mreža podataka	357
Analitika	358
Poslovna analitika	358
Operativna analitika	360
Ugrađena analitika	363
Mašinsko učenje	364
Šta inženjer podataka treba da zna o ML	364
Načini serviranja podataka za analitiku i ML	366
Razmena datoteka	366
Baze podataka	367
Sistemi strimovanja	368
Federacija upita	369
Deljenje podataka	370
Semantički i metrički slojevi	370
Serviranje podataka u beležnicama	371
Reverzan ETL	373
Sa kime ćete raditi	375
Podtokovi	375
Bezbednost	376
Upravljanje podacima	377
DataOps	377
Arhitektura podataka	378
Orkestracija	378
Inženjerstvo softvera	379
Zaključak	380
Dodatni izvori	381

---

## Deo III Bezbednost, privatnost i budućnost inženjerstva podataka

<b>10. Bezbednost i privatnost.....</b>	<b>385</b>
Ljudi	386
Moć negativnog mišljenja	386
Uvek budite oprezni	387
Procesi	387
Pozorište bezbednosti u odnosu na bezbednosnu naviku	387
Aktivna bezbednost	388
Princip najmanjih privilegija	388
Deljena odgovornost u oblaku	388
Uvek napravite rezervnu kopiju svojih podataka	389
Primer bezbednosne politike	389



Tehnologija	390
Ažuriranje i ispravke sistema	391
Šifrovanje	391
Prijavljivanje, praćenje i upozorenja	392
Pristup mreži	393
Bezbednost za nizak nivo inženjerstva podataka	393
Zaključak	394
Dodatni izvori	394
<b>11. Budućnost inženjerstva podataka .....</b>	<b>395</b>
Životni ciklus inženjerstva podataka neće nestati	396
Opadanje složenosti i porast lako upotrebljivih alata za podatke	396
Klaud skalabilni OS za podatke i unapređena interoperabilnost	397
Preduzeću slično inženjerstvo podataka	399
Pozicije i odgovornosti će se menjati...	400
Izlazak iz okvira savremenog steka podataka, ka živom steku podataka	401
Data Stack uživo	402
Strimovanje cevovoda i baze podataka za analitiku u realnom vremenu	402
Fuzija podataka sa aplikacijama	404
Čvrsta povratna informacija između aplikacija i ML	404
Tamna strana podataka i uspon...tabela?!	404
Zaključak	405
<b>A. Serijalizacija i kompresija: tehnički detalji.....</b>	<b>407</b>
<b>B. Umrežavanje u oblaku.....</b>	<b>415</b>
<b>Indeks .....</b>	<b>419</b>