

Tehnološke preporuke na sustavima CEKOM

PROJEKTNI ZAHTJEVI

.

Sadržaj

1.	TEHNOLOŠKE PREPORUKE NA SUSTAVIMA CEKOM	3
1.1.	Zajednička horizontalna infrastruktura sustava CEKOM	3
1.2.	Preporuke arhitekture	3
	Ciljevi i opseg sustava	3
	Arhitektura sustava	4
	Planiranje kapaciteta	4
	Implementacijske preporuke	6
1.3.	Ostale Digitalne tehnologije	6
	Analiza mogućnosti korištenja digitalnih tehnologija gledana iz perspektive EU	7
	Trenutni projekti i otvorene mogućnosti investiranja na razini EU	7
	Analiza mogućnosti proširenja portfelja podatkovnih proizvoda i usluga gledana iz perspektive hrvatskog strateškog plana razvoja infrastrukture	7
	Podatkovna zanosnost (data science) i strojno učenje (machine learning)	8
2.	LITERATURA (za zahtjeve)	9

1. TEHNOLOŠKE PREPORUKE NA SUSTAVIMA CEKOM

1.1. Zajednička horizontalna infrastruktura sustava CEKOM

CEKOM ima 8 kolaborativnih aktivnosti:

- PA 1: "Smart Agri" - Istraživanje modela i razvoj platforme za primjenu IoT / M2M u poljoprivredi na ruralnim područjima
- PA 2: SW platforma za potporu odlučivanju u upravljanju adaptibilnošću poslovnih sustava (zadruga, MSP, OPG, LAG, 5-Helix)
- PA 3: Distribucijska platforma za prehrambene proizvode po modelu „Platform based business“
- PA 4: Razvoj SW i portala za kolaboraciju krajnjih korisnika u području poljoprivrede i prehrambene industrije na ruralnim područjima
- PA 5: SW i portal za samokontrole za sanitarne regulatorne okvire u proizvodnji hrane
- PA 6: "Smart Food" radni okvir, standardi i primjena robotike u proizvodnji hrane
- PA 7: "Smart Farm" primjena za zaštitu od bolesti životinja i od nepogoda
- PA 8: Digitalna platforma za plasman hrane u zdravstvenom turizmu kao distribucijski kanal

Po zajedničkim horizontalnim infrastrukturnim elementima interno definiranih kako slijedi:

HI 1	AI & BigData & Knowledge management
HI 2	IoT & DL & DLT
HI 3	Digitalni ekosustavi & virtualne zajednice
HI 4	Digitalne platforme & Platform Based Business
HI 5	Cyber Security
HI 6	Interoperabilnost i integracijski servisi
HI 7	Suradnja u ekosustavima za međunarodnu suradnju i Izvoz
HI 8	Upravljanje ukupnim životnim ciklusom proizvoda - DevOps
HI 9	BPM & Flowable
HI 10	EA & Alati za modeliranje
HI 11	Open Data & MDM & Javne baze znanja
HI 12	EU infrastrukturni elementi - CEF DSI Building Blocks & Fiware
HI 13	Platforma za Ankete (EUSurvey)
HI 14	Portalske tehnologije

Dobiveni rezultati trebaju se moći koristiti kroz sve projektne aktivnosti.

1.2. Preporuke arhitekture

Ciljevi i opseg sustava

Cilj je sustava pružiti mogućnosti spremanja, analize, prezentacije i razmjene (interoperabilnost) podataka vezanih uz prerade, pakiranja hrane i istraživanja.

Predlaže se da arhitektura podržava sljedeće funkcionalnosti:

1. Unos podataka u „batch“ i „streaming“ načinu
2. Pohrana podataka vrlo velikih skupova podataka
3. Serijska obrada
4. Stream obrada

5. Pretraživanje
6. Eksplorativna analiza podataka, izvještavanje/poslovna inteligencija i primjena metoda umjetne inteligencije
7. Prediktivno modeliranje

Arhitektura sustava

Arhitektura sustava temeljila bi se na referentnoj arhitekturi velikih podataka (primjer tvrtke Cloudera Inc. koja preporuča/koristi ili slično sljedeće komponente:

1. Apache Sqoop za unošenje serijskih podataka
2. Apache Kafka za unos podataka u streaming
3. MapReduce za „batch“ obradu
4. Apache Spark za „batch“ obradu, „streaming“ obradu i prediktivno modeliranje
5. Cloudera Search funkcionalnost pretraživanja
6. Apache Impala za eksplorativnu analizu podataka, izvještavanje/poslovnu inteligenciju i primjenu metoda umjetne inteligencije
7. Server i radna površina za istraživačku analizu podataka, izvještavanje i vizualizaciju (Tableau, PowerBI, Alteryx ili sličan)

Planiranje kapaciteta

Pohrana

Pohranu u HDFS-u preporučeno je konfigurirati s faktorom replikacije 3, time su zahtjevi za pohranom tri puta veći od stvarnih zahtjeva za podacima. Za spremanje privremenih rezultata potreban je dodatni prostor na disku. Trenutna industrijska praksa je da se diskovi drže na 70% kapaciteta. Na temelju trenutno dostupnih informacija o izvorima podataka i uzimajući u obzir izvore podataka koji bi se mogli dodati u sustav u dogledno vrijeme, potreban kapacitet pohrane podataka procjenjuje se na 8 TB. Pohrana podataka treba biti lako skalabilna dodavanjem više diskova postojećim podatkovnim čvorovima ili dodavanjem novih podatkovnih čvorova u klaster.

Obrada

Procjenjuje se da je procesorska snaga potrebna za unos podataka, skupnu obradu, obradu „stream“-a i strojno učenje zadovoljna sljedećom referentnom konfiguracijom: (i) broj čvorova u klasteru četiri (4); (ii) referentna vrsta poslužitelja: HPE ProLiant DL380 GenX ili veći; (iii) iip procesora Intel Xeon E5-2650 v4 – najmanje dva (2) fizička procesora po poslužitelju

Memorija

Budući da se očekuje da će se sustav koristiti za skupnu obradu podataka, kao i za interaktivnu analizu, potrebna memorija procjenjuje se na minimalno 32 GB po čvoru. Memorija treba također biti lako proširiva s kapacitetom do 3TB po poslužitelju.

Umrežavanje

Poslužitelji su opremljeni 1Gb Ethernet kontrolerom s 4 porta po kontroleru. Smatra se da je 16-portni 1Gb preklopnik za poduzeće dovoljan za procijenjeno radno opterećenje.

Programska podrška

Sustav se temelji na industrijskom standardnom paketu velikih podataka s većinom rješenja otvorenog koda spakiranih u komercijalnu distribuciju. Preporučeni operativni sustav je Linux korporativnog razreda, kao što je Red Hat Enterprise Linux. Proizvodi za vizualizaciju koriste se za jednostavan i intuitivan pristup podacima, upite, vizualizaciju i istraživačku analizu.

HPC infrastruktura

Za potrebe složenih simulacija, preporučeno je dugoročno planirati HPC kapacitete pri čemu treba voditi računa o brzom razvoju ovih tehnologija (planirati sustave na arhitekturama koje je moguće jednostavno skalirati) te potrebama kontejnerizacije podataka. Preporuka je voditi se OpenStack arhitekturom te vScaler Appliance rješenjima tvrtke Boston Consulting. Ista

AI podrška

Kvalitetna osnova arhitekture uz omogućenu podršku nadogradnji HPC sustavima omogućava primjerene podloge za primjenu AI metoda u analizi podataka i konačnim rješenjima/uslugama baziranim na podacima.

Implementacijske preporuke

Uz infrastrukturu, važna komponenta je planiranje ukupnih troškova odnosno ukupnih troškova implementacije. Pri tome konačno rješenje mora sadržavati procjenu resursa koji su povezani su sa:

1. Priprema okoline
2. Instalacijom i konfiguracijom sustava
3. Preliminarnom analizom podataka
4. Unosom podataka
5. Dizajnom podatkovnog sloja
6. Obradom podataka (transformacija, agregiranje, itd.)
7. Prediktivnim modeliranjem
8. Provjerom valjanosti modela
9. Izradom podatkovnih modela (za svaki predloženi model)
10. Izvještavanje, nadzorna ploča, upozoravanje i druge BI funkcionalnosti
11. Projektna dokumentacija
12. Obuka krajnjih korisnika
13. Održavanje sustava
14. Unapređenja sustava
15. (Integracije sa vanjskih sustavima)
16. (Čišćenje podataka)

Implementacija rješenja temelji se na informacijama dostupnim u vrijeme pisanja ovog dokumenta. I arhitektura i procjena troškova daju se pod pretpostavkom da su potrebni podaci dostupni i dovoljno kvalitetni odnosno u slučaju da postoje izazovi sa kvalitetom podataka, u procjenu je potrebno uključiti troškove integracije vanjskih sustava ili čišćenja podataka.

1.3. Ostale Digitalne tehnologije

Pristup razumijevanju potencijala digitalnih tehnologija a posebno podatkovne znanosti u ovom području je promatrati aktivnostima kojima se bave. Postoje tri skupa aktivnosti koje se ponavljaju a važne su za ovo područje:

1. Preobrazba poslovnih izazova u analitičke – dijagnosticiranje poslovnih problema, razmatranje srži problem a i naposljetku zaključivanje koje se analitičke metode mogu koristiti za rješavanje tih problema
2. Dizajniranje, implementacija i postavljanje statističkih metoda, te tehnike rudarenja podataka (eng. data mining) nad velikim podacima – ovo su aktivnosti na koje se najčešće pomisli kada razmatramo ulogu podatkovne znanosti. Korištenje kompleksne ili napredne analitičke metode za razne poslovne probleme koristeći podatke
3. Stvaranje uvida koji vode do prijedloga aktivnosti – korištenje naprednih metoda nad podatkovnim problemima ne mora donositi nove poslovne vrijednosti, nego je bitno naučiti kako izvlačiti uvide iz podataka i efikasno ih komunicirati.

Analiza mogućnosti korištenja digitalnih tehnologija gledana iz perspektive EU

Važnost investiranja u ICT sektor se najbolje vidi iz statistika EU trenutnih i budućih investicija koja jasno govore u kojem smjeru se dominantno razvija generalna slika investicija analiziranih preko svih strateških područja. Kroz razne smjernice i agende od kojih valja izdvojiti Digital Agenda for Europe¹, Digital Single Market kao dio Europe 2020 strategije² i među najnovijim S3 Smart Specialization Platform³ koja uključuje investiciju od preko 500 mil. EUR za podršku preko 500 Digital Innovation HUB lokacija u EU odnosno njen produžetak koji tijekom 2021. očekuje formiranje mreže EDIH lokacija. Upravo propagacijom sredstava kroz takve aktivatore koji kombiniraju razne ekspertize otvaraju se velike mogućnosti.

Trenutni projekti i otvorene mogućnosti investiranja na razini EU

Ključni projekti i otvorene mogućnosti u koje trenutačno najviše investira na razini EU se odnose upravo na ICT usluge i infrastrukturu međutim ne kao horizontalne teme već kao teme koje su izravno povezane sa nekim od područja, u konkretnom slučaju – „pametnom hranom“⁴. U prosjeku se u ICT sektor investira 5 do 10 puta više nego u bilo koji drugi sektor⁵ pri čemu se dalji rast i otvaranje raznih oblika financiranja za podatkovne/AI i srodne proizvode i usluge bazirane na digitalnim tehnologijama očekuju u mjesecima koji dolaze.

Važna područja investiranja će biti usmjerena na:

1. Održavanje i razvoj informacijskih sustava,
2. Konzultantske usluge za IT sustave Europske unije i primjene u raznim područjima e-uprave,
3. Usluge razvoja životnog ciklusa razvoja informacijskih sustava,
4. Ulaganje u kibernetičku sigurnost, AI, HPC i digitalne vještine

Promatrajući navedene trendove a imajući u vidu da se kod podatkovnih proizvoda i usluga radi o visokoj dodanoj vrijednosti, relativno dugom prodajnom ciklusu te da ih prati karakteristika razvoja i istraživanja u pri čemu je kontinuirano potrebno ulagati u resurse i znanja, očekuje se rast potražnje i već danas je očit manjak konkretnih rješenja uz višestruko povećanje ulaganja u znanja.

Analiza mogućnosti proširenja portfelja podatkovnih proizvoda i usluga gledana iz perspektive hrvatskog strateškog plana razvoja infrastrukture

Ministarstvo znanosti, obrazovanja i sporta je prije nekog vremena objavilo još uvijek aktualan „Plan razvoja istraživačke infrastrukture u Republici Hrvatskoj“ koje među ostalim navodi da je ulaganje u ICT, kao horizontalnim sektorom o kojem ovise gotovo svi ostali sektori, pitanje od najvećeg strateškog prioriteta. I kao takvo predstavlja ex-ante uvjet za ulaganja u istraživanje, razvoj i inovacije sukladno utvrđenoj metodologiji. Glavni strateški cilj Hrvatske je transformacije

¹ http://eige.europa.eu/resources/digital_agenda_en.pdf

² <https://ec.europa.eu/digital-single-market/en/europe-2020-strategy>

³ <http://s3platform.jrc.ec.europa.eu/>

⁴ https://ec.europa.eu/food/farm2fork_en

⁵ http://www.c4p.io/wp-content/uploads/2019/06/C4P-Newsletter_June19.pdf

hrvatskog gospodarstva kroz učinkovite aktivnosti istraživanja, razvoja i inovacija što je nemoguće bez kvalitetne strukture investiranja u ICT sektor.

Jedan od ključnih ciljeva koji se pri tome navode je partnerski rad na rješavanju društvenih izazova pri čemu kvalitetnija realizacija različitih usluga pojedinih ministarstava igra dominantnu ulogu. Pri tome se poticanje suradnje i protoka znanja između poslovnog, javnog i znanstveno-istraživačkog sektora navodi kroz jedna od temeljnih tematskih stupova Strategije.

Od izuzetne je važnosti prepoznati potencijali i snage za jačanje područja ICT povezano s prepoznatim tematskim prioritetnim područjima definiranim u strategiji Republike Hrvatske i pod područjima Strategije pametne specijalizacije su koja se odnose na ICT. Konkretna pod područja prepoznata kao dominantna među ostalim su velike količine podataka (eng. Big Data) i usluge temeljene na internetu.

Istraživačko-razvojne vezane uz Mreže budućnosti (eng. Future Networks) i Internet budućnosti (eng. Future Internet) pokrivaju, prema područjima primjene definiranim kroz TPA, tri povezana dijela globalnog lanca vrijednosti: Internet stvari (komunikacijski softver i platforme za međusobno povezane objekte), velike podatkovne baze eng. Big Data (pribavljanje, obrada i analiza podataka iz fizičkog i virtualnog svijeta) i usluge temeljene na internetu, a primjene su moguće u gotovo u svim prioritetnim područjima.

Podatkovna zanosnost (data science) i strojno učenje (machine learning)

Ključne novosti iz raznih područja EU investicija idu u prilog području podatkovne znanosti i strojnog učenja koje omogućava bolje upravljanje podacima, uslugama i mase ostalih pogodnosti koje olakšavaju kvalitetniji razvoj e-uprave, poslovanju u globalu te jednostavniji i bolji protok kvalitetnih informacija unutar informacijskog sustava države.

Europski naponi u implementaciji podatkovnih projekata očituju se i kroz forume pri čemu je EDF ključni europski događaj za industrijske stručnjake, poslovne partnere, istraživače i kreatore politike kako bi razmotrili izazove i mogućnosti europskog gospodarstva podataka i inovacije temeljene na podacima u Europi. Ove godine EDF i BDVA udružuju snage Europski forum o velikim podacima (European Big Data Value Forum⁶).

Ključna područja projekta i investiranje u području podatkovne znanosti, strojnog učenja i umjetne inteligencije tematski su vezana uz područje "Pouzdana umjetna inteligencija u pametnoj industriji". Iz napora EU u ovom području koje se očituje kroz izuzetno visoke razine investicija, velik broj različitih projekata i aktivne implementacije u sve vidove gospodarstva i javne uprave, očigledna je EU vizija društva utemeljenog na podatkovnoj znanosti. Projekti koji su dominantni u području se baziraju na analize podataka u osobnim, profesionalnim i industrijskim kontekstima te se sve više i više digitalnih usluga implementira na temelju istih. Automatizirano odlučivanje temeljeno na podacima, a bazirano na sofisticiranim algoritmima aktivno se implementira u različitim domenama od obrade osjetljivih osobnih podataka do kritičnih industrijskih ugrađenih sustava. Transparentnost i odgovornost takvih algoritamskih sustava temeljenih na podacima postaju sve važnija za povjerenje i pouzdanost digitalnih usluga u B2C i B2B kontekstu. Oni su faktori konkurentnosti za inovacije utemeljene na podacima.

Europski forum za Big Data je ujedno platforma za "veliko javno-privatno partnerstvo" u području Big Data koju su uspostavili Europska komisija i europski dionici s podacima i samim time je idealan putokaz za Hrvatsku u pogledu smjernica daljnjeg razvoja projekata i investicija baziranih na podatkovnim tehnologijama pri čemu se aktivno objedinjavaju naponi javne uprave, raznih ministarstava, istraživačkih centara te različitih gospodarskih subjekata. Takovi oblici međusobne suradnje su uvelike podržani te se sve intenzivnije provode na razini EU.

⁶ <http://www.european-big-data-value-forum.eu/>

Neke od ostalih područja vezanih za podatkovnu znanost u koje se aktivno ulaže su:

inovacije bazirane na podacima, transport, informacijskih sustavi unaprijeđeni sa strojnim učenjem . Značajan razvoj u tom području na razini EU postignut je sa pristupom sve kvalitetnijim podacima; pri čemu se poradi toga pojavljuju novi alati i radna okruženja koja podržavaju razvoj sve izvrsnijih algoritama koji pak omogućavaju razvoj cijelog niza alata⁷ u području govora, civilnog društva, akademske zajednice, novinarstva , industrije, managementa ljudskih resursa, transporta, lanca opskrbe, itd.

2. LITERATURA (za zahtjeve)

Literatura:

- Plan razvoja istraživačke infrastrukture u Republici Hrvatskoj
<https://mzo.gov.hr/UserDocsImages//dokumenti/Znanost/TransferTehnologije//Plan%20razvoja%20istra%C5%BEiva%C4%8Dke%20infrastrukture%20u%20Republici%20Hrvatskoj%20-%20lipanj%202016.pdf>
- Farm to Fork Strategy – for a fair, healthy and environmentally-friendly food system
https://ec.europa.eu/food/farm2fork_en
Shaping a Climate-Smart Global Food System
<https://www.worldbank.org/en/topic/agriculture/publication/shaping-a-climate-smart-global-food-system>; <https://openknowledge.worldbank.org/handle/10986/22927>

⁷ <http://www.european-big-data-value-forum.eu/program/>