# WHY: Real-life

Intregul proiect care sta la baza acestei teze a plecat de la o nevoie reala determinata in cadrul unui proiect comercial ~~si a proceselor de cercetare si dezvoltare experimentala realizate in cadrul acestui proiect~~. Mai concret, problema identificata a fost aceea de a executa simultan un ansamblu de experimente big-data fara a utiliza infrastructura externa (cloud, cluster).

~~Cele 2 elemente de baza ale noastre sunt arhitecturile grafurilor si volumul mare de date.~~ Referitor la executia grafurilor computationale, cunoastem ca acestea pot avea aplicabilitati in diverse domenii precum: cryptologie, fizica, procesare video etc. In particular, executia, dar si optimizarea grafurilor computationale tensoriale in domeniul Analizei Predictive (Predictive Analytics) este anevoioasa atunci cand trebuie procesate masive de date prin experimente multiple cu responsivitate maxima. Este evident ca aceste lucruri se pot realiza folosind procesare in cloud sau pe cluster, insa in acest caz intervine problema confidentialitatii datelor. Voi da un exemplu foarte scurt: se doreste realizarea unui “Proof-of-Concept” la un client ale carui masive de date pot fi accesate numai prin VPN local. Cum facem sa i le procesam “on-the-spot” din perspective multiple, intr-un timp rezonabil, fara a pune in pericol sensibilitatea datelor (mai exact “a plecarii” cu datele din locatia respectiva)? Astfel ajungem la motivatia tehnologica a proiectului.

# WHY: Technical

Plecand de la nevoia reala identificata, vom lua cazul executarii paralele a experimentelor noastre bazate pe grafuri si big-data intr-un mediu computational ce dispune de capabilitati de procesare paralela numerica, desi acestea sunt limitate, cum ar fi cazul unui laptop performant, sa zicem un notebook cu o placa grafica de NVIDIA cu aprox. 1000 coreuri numerice si 6GB memorie video. Si mai concret de atat, scopul nostru dpdv tehnic este sa gasim o solutie de a analiza grafurile, de a det. Max de resurse necesare pt executia acestora si de a le executa intr-un format „Best-Fit” (ocupand la max. Resursele disponibile).

# Related Work: TensorFlow

Analizand specificatiile de nivel inalt ale motivatiei tehnice, ajungem in punctul in care trebuie sa definim tehnologiile pe care se bazeaza proiectul. Tehnologiile folosite de noi se regasesc indubitabil in stadiul actual al cercetarii si tehnologiei si in acest sens ne vom referi doar la framework-ul de procesare si executie a grafurilor tensoriale in medii hibride – TensorFlow, lansat de Google in 2016.

Dpdv al paralelizarii calculelor numerice, framework-ul TensorFlow ne ofera posibilitatea de a abstractiza scrierea kernel-elor NVIDIA CUDA. Cu toate acestea, acest framework considerat actualmente cel mai important si cel mai folosit la nivel international, nu a fost proiectat si nu are capabilitati de a oferi suport nativ pentru problema noastra tehnica.

Ca interfata de programare, mentionez ca in intreg proiectul se foloseste limbajul Python.

# Related Work: P2V

Acest model grafic pe care vi-l prezint in continuarea enumerarii elementelor de baza ale tezei, este rezultatul cercetarii realizate de noi cuprinsa in cadrul lucrarii stiintifice *Deep recommender engine based on efficient product embeddings neural pipeline*. Practic, ceea ce vedeti este o harta vizuala a intregului ecosistem de produse/servicii ale utilizatorului real al acestui proiect, in particular o companie de farmaceutice. Fara a intra in detalii, mentionez ca aceasta harta defineste relatii de complementaritate si similaritate intre oricare 2 entitati ale ecosistemului. Putem spune ca rezultatele cercetarii si dezvoltarii acestui proiect si prezentate in teza de fata, au avut ca scop facilitarea obtinerii acestei harti in locatie si in timp minim. Mentionez de asemenea, ca pe langa aceasta harta, mai exista si alte obiective si rezultate similare, care nefacand scopul prezentei teze, pot fi regasite in cadrul lucrarii stiintifice.