

Dan slovenskega superračunalniškega omrežja, MREŽA ZNANJA
16. november 2023

arnes 



EURO



SLING

Computer Vision in Natural Language Processing od ideje do superračunalnika

Prezenter: Nejc Bat, XLAB d.o.o.

XLAB
**Software development
podjetje**

2001
**Ustanovljeno v TP,
Ljubljana**

100+
zaposlenih

EXPERTIZE
**IT Automation &
Management**

AI & Machine Learning

App Modernization





CYLCOMED

Cybersecurity Toolbox for Connected Medical Devices



AIDEAS

AI-driven industrial equipment product life cycle



SUNRISE

Resilient Critical Infrastructures and Vital Services



TANGO

Digital technologies for secure data flows



ICOS

Functional continuum operating system



CNCSmart

AI-based machining time prediction



SODALITE

Application & Ops Performance Management



RADON

Emerging Serverless Computing Technologies



PIACERE

Trustworthy IaC in DevOps integrated framework



FISHY

Cyber Resilience Provisioning Framework



MEDINA

Security framework for continuous cloud certification



ARCADIAN-IoT

Making IoT systems secure



KRAKEN

Privacy Preserving Platform



PIXEL

Port IoT Environmental Leverage



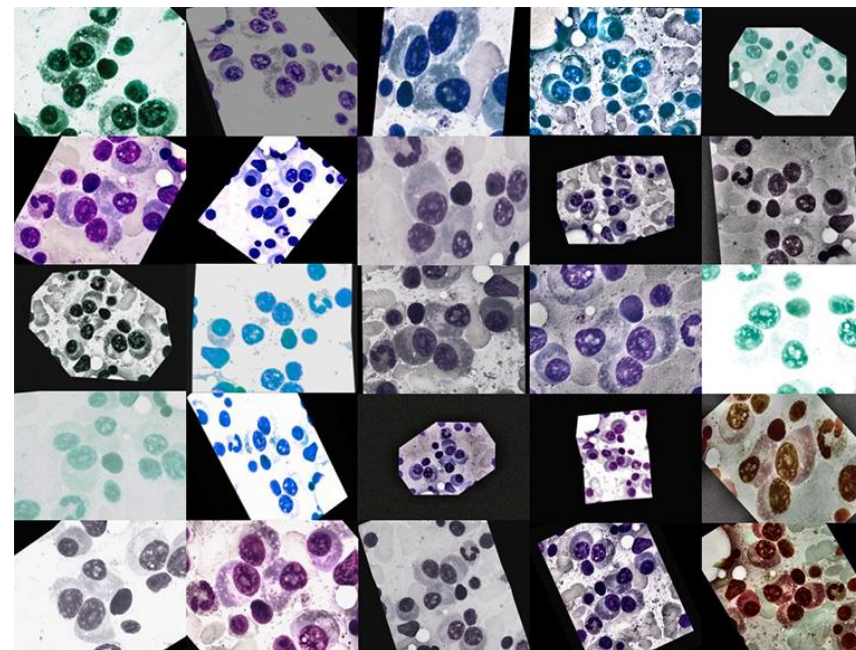
iPC

Individual Paediatric Cure

Kaj počnemo s CV in NLP?



- Prepoznavna vzorcev v slikah rakavih celic
- Prepoznavna napak v materialu (pred proizvodnjo) in končnih proizvodih (anomalije)
- Iskanje anomalij v log-ih
- Veliki jezikovni modeli



Kje se začnejo HPC izzivi?



- Količine podatkov se večajo
- Modeli se večajo
- Algoritmi postajajo kompleksnejši



Kako jih sprva rešimo?



- Optimizacija kode
- Overnight procesiranje
- Boljši PCji
- PCji v mrežo



Ali je to trajna rešitev?



- Ekipa AI ekspertov se širi
- Količine podatkov se večajo
- Modeli se večajo
- Algoritmi postajajo kompleksnejši



Kakšna je naslednja rešitev?



- Kupimo lasten strežnik



Ali je to trajna rešitev?



- Ekipa AI ekspertov se širi
- Količine podatkov se večajo
- Modeli se večajo
- Algoritmi postajajo kompleksnejši
- Strežnik mora v kolokacijo
- Administrirati moraš strežnik – nove kompetence
- Strežnik postane nezanesljiv in nezadosten za potrebe

Kaj se zgodi potem?



- Al ekipa predlaga, da bi kupili še en novejši strežnik
- Direktor ne odobri nakupa
- Končno izračunamo TCO
 - Nakup HW
 - Vzpostavitev strežnika
 - Čas vzdrževanja
 - Električna
 - Utilizacija
- Kakšne so naše dejanske potrebe?

Kakšne so naše potrebe?



- GPU:
 - Več GPU-RAM je bolje
 - Trenutno najzahtevneši primeri tečejo na 16 GPUjih hkrati
- Podatkovna hramba
 - 100 GB+ veliki data set za posamezni primer
 - Skupno trenutno obdelujemo 30 TB podatkov
- Konteinerizacija

Zakaj VEGA



- Večina inženirjev ima izkušnje s SLING kot študenti
- Primerni resursi (GPU-RAM)
- Support, lokalni vidik
- TCO = 0€



Hvala!

Projekt EuroCC 2 financira Evropska unija. Financiran je s sredstvi Skupnega evropskega podjetja za visokozmogljivo računalništvo (EuroHPC JU) ter Nemčije, Bolgarije, Avstrije, Hrvaške, Cipra, Češke republike, Danske, Estonije, Finske, Grčije, Madžarske, Irske, Italije, Litve, Latvije, Poljske, Portugalske, Romunije, Slovenije, Španije, Švedske, Francije, Nizozemske, Belgije, Luksemburga, Slovaške, Norveške, Turčije, Republike Severne Makedonije, Islandije, Črne gore in Srbije v okviru sporazuma o dodelitvi sredstev št. 101101903.



**Funded by
the European Union**