

# Паралелно и дистрибуирано процесирање

Домашна задача 1

Петар Атанасовски - 216052

## 1.1.

### a. High-performance computing (HPC) system:

HPC користи паралелно процесирање и напредни алгоритми за решавање на сложени проблеми, честопати се потпира на повеќе десктоп компјутери или сервери како и на моќен хардвер како суперкомпјутери.

(1.1.1.2, pg. 5)

### b. High-throughput computing (HTC) system:

HTC се фокусира на ефикасна обработка на голем број задачи за релативно кратко време. За разлика од HPC, кој има за цел процесирање на една задача со голема брзина, HTC има цел да се справи со многу задачи истовремено.

### c. Peer-to-peer (P2P) network:

P2P е децентрализиран модел каде корисниците имаат еднаков статус и можат директно да комуницираат или споделуваат ресурси, за разлика од моделите клиент-сервер каде што серверот обезбедува услуги на клиенти.

### d. Computer cluster versus computational grid:

Computer cluster е група на меѓусебно поврзани компјутери кои соработуваат преку споделување процесорска моќ и ресурси за да постигнат заедничка цел. Computational grid го проширува овој концепт со поврзување на ресурсите за уште поефикасна обработка на податоци.

### e. Service-oriented architecture (SOA):

SOA е архитектонски пристап кај кој имаме лабаво поврзани софтверски компоненти кои комуницираат преку мрежа, овозможувајќи модуларност и повторна употреба за создавање поголеми апликации.

### f. Pervasive computing versus Internet computing:

Pervasive computing ги интегрира компјутерите во секојдневниот живот и околини за нивна интеракција со луѓето, додека Internet computing се фокусира на доставување пристап до апликации и услуги преку интернет.

### g. Virtual machine versus virtual infrastructure:

Virtual machine е софтверска имитација на физички компјутер на кој може да има оперативен систем и апликации. Virtual infrastructure ја опфаќа целата околина што поддржува VMs, вклучувајќи го основниот физички хардвер, софтверот за виртуелизација и алатките за управување.

### h. Public cloud versus private cloud:

Public cloud е услуга понудена од провајдери, обезбедувајќи ресурси за јавноста, додека private cloud е инфраструктура управувана исклучиво за една организација, која нуди поголема контрола и безбедност.

**i. Radio-frequency identifier (RFID):**

Radio-frequency identifier (RFID)-технологијата користи радио бранови за идентификација и следење на објект. Објектот има мал чип, а читач комуницира со него и добива информации емитувајќи радио бранови.

**j. Global positioning system (GPS):**

GPS е сателитски навигациски систем кој широко се користи за обезбедување информации за локацијата и времето во сите временски услови. Се користи во возила, телефони и различни апликации.

**k. Sensor network:**

Sensor network е збир на дистрибуирани автономни сензори кои ги следат физичките или еколошките услови, комуницирајќи едни со други и потенцијално испраќајќи податоци до централна локација за обработка и анализа.

**l. Internet of Things (IoT):**

IoT е меѓусебно поврзување на секојдневните физички објекти на интернет, овозможувајќи им да испраќаат и примаат податоци и нуди далечинско следење, контрола и интеракција.

**m. Cyber-physical system (CPS):**

CPS интегрира компјутерски елементи (компјутери, софтвер) со физички компоненти (сензори) за да создаде системи кои комуницираат и влијаат на физичкиот свет. Од клучно значење е за апликации со автоматизација како паметни градови и автономни возила.

**1.2.**

**1. In the 2009 Top 500 list of the fastest computer systems, which architecture dominates?**

b. Centralized massively parallel processor (MPP) systems.

**2. In a cloud formed by a cluster of servers, all servers must be selected as follows:**

c. The cloud machines can be either physical or virtual servers.

**1.3.**

**a. List and describe the main characteristics of cloud computing systems.**

- ☐ **On-Demand Self-Service:** Корисниците можат да управуваат со компјутерските ресурси (сервери, складирање и вмрежување) по потреба, без дополнителна интеракција со друг service provider.
- ☐ **Broad Network Access:** Cloud услугите се достапни преку интернет на различни уреди, како што се лаптопи, паметни телефони и таблети.
- ☐ **Resource Pooling:** Cloud provider-ите динамично доделуваат ресурси на повеќе корисници врз основа на побарувачката, за физички и за виртуелни ресурси. Корисникот не знае од која точна локација ја има добиено соодветната информација.

- ☐ **Rapid Elasticity:** Cloud ресурсите се лесно скалабилни и брзо се приспособуваат на промените на побарувачката. Ова овозможува флексибилност и економичност.
- ☐ **Measured Service:** Cloud системите користат автоматска контрола на ресурсите и оптимизација врз основа на побарувачката на ресурси.

**b. Discuss key enabling technologies in cloud computing systems.**

- ☐ **Virtualization:** Оваа технологија овозможува повеќе виртуелни машини (VM) или контејнери да работат на една физичка машина
- ☐ **Distributed Computing:** Им овозможуваат на cloud системите да обработуваат задачи преку повеќе сервери, подобрувајќи ги перформансите преку паралелна обработка.
- ☐ **High-Speed Networking Technologies:** Обезбедуваат беспрекорна комуникација меѓу компонентите на облакот.
- ☐ **Storage Technologies:** ефикасно управуваат со големи количини на податоци, тука се вклучуваат NAS, SAN и складирање на објекти.
- ☐ **Security and Encryption:** Спаѓаат шифрирање, заштитен сид, контроли за пристап и автентикација, заштита на податоци и апликации.
- ☐ **Orchestration and Management Tools:** Го автоматизираат распоредувањето на ресурсите, скалирањето и координацијата во cloud средини.

**c. Ways for Cloud Service Providers to Maximize Revenues:**

- ☐ **Subscription-Based Models:** Ововможуваат редовен пристап заснован на претплата (супскрипција).
- ☐ **Pay-Per-Use Pricing:** Наплаќање врз основа на реалната употреба.
- ☐ **Value-Added Services:** Обезбедување дополнителни услуги како анализа на податоци и напредна безбедност за повисоки цени.
- ☐ **Customized Solutions:** Прилагодување на понуди за да се задоволат специфичните потреби на клиентот.
- ☐ **Partnerships and Ecosystem Development:** Соработка за создавање интегрирани решенија и можности за споделување приходи.
- ☐ **Geographic Expansion:** Проширување на услугите во нови региони или земји.
- ☐ **Continuous Innovation:** Да се биде во тек со новата технологијата со цел задржување и привлекување клиенти.
- ☐ **Transparent Pricing and SLAs:** Правење цврсти договори за да се изгради доверба и да се обезбедат долгорочни партнерства.

**1.4.**

**Match the terms and systems:**

Globus	H
BitTorrent	F
EC2	I
TeraGrid	D
EGEE	C
Hadoop	A
SETI@home	J

Napster	B
BigTable	E
Gnutella	G

### 1.5.

$$\text{Total execution time A} = \max\left(\frac{32}{1}, \frac{128}{2}, \frac{64}{3}, \frac{32}{1}\right) = 64(\text{unit time})$$

$$\text{Total execution time B} = \max\left(\frac{48}{1}, \frac{128}{2}, \frac{80}{3}, \frac{0}{1}\right) = 64(\text{unit time})$$

$$\text{Utilization A} = \frac{\frac{32}{1} + \frac{128}{2} + \frac{64}{3} + \frac{32}{1}}{4} \times \frac{1}{64} \approx 0.58$$

$$\text{Utilization B} = \frac{\frac{48}{1} + \frac{128}{2} + \frac{80}{3}}{3} \times \frac{1}{64} \approx 0.54$$

### 1.8.

#### a. Hardware, Software, and Networking Support:

##### ☐ Traditional Clusters/Grids:

- **Hardware:** Хомогени сервери поврзани преку мрежа со голема брзина, често прилагодени за специфични апликации.
- **Software:** Специјализиран софтвер за управување со кластери/мрежа за распоред на задачи и распределба на ресурси.
- **Networking:** Потребни се интерконекции со високи перформанси и ниска латентност за ефикасна комуникација.

**Пример:** Beowulf Cluster - се користи за научни и инженерски пресметки.

##### ☐ Recent Computing Clouds:

- **Hardware:** Хетерогени сервери, често овозможена виртуелизација, дизајнирани за приспособливост и флексибилност.
- **Software:** Користи виртуелизација и контејнеризација (containerization) за ефикасност на ресурсите.
- **Networking:** Се потпира на брзи интернет конекции и виртуелни мрежи за интерконекција со ресурси.

**Пример:** Amazon Web Services (AWS) - нуди широк опсег на виртуелизирани услуги за различни апликации.

#### b. Resource Allocation and Provisioning Methods:

##### ☐ Traditional Clusters/Grids:

- **Resource allocation** е статична и поставена за време на поставувањето на кластерот.
- **Scaling** вклучува додавање физички јазли или надградба на постоечките.

##### ☐ Recent Computing Clouds:

- **Dynamic resource allocation** овозможува скалирање, овозможувајќи обезбедување или одземање на барање.

- **Auto-scaling** ги прилагодува ресурсите врз основа на барањата за обемот на работа.
- c. Infrastructure management and protection**
  - Cloud computing се потпира на виртуелизација на сите видови ресурси и му треба посилна заштита од кластерот/мрежата.
- d. Support of utility computing services**
  - Cloud computing обезбедува комунални компјутерски услуги, слични на електрична енергија и вода, додека кластерите/мрежите обично се распоредени од поединечни претпријатија и не се нудат како јавни.
- e. Operational and cost models applied**
  - Cloud computing е генерално поисплатливо од cluster/grid setups бидејќи клиентите не мора однапред да инвестираат во инфраструктура. Договорите за квалитет како услуга (QaaS) и договори за ниво на услуга (SLA) наметнуваат дополнителни барања за оперативните и моделите на трошоци.

#### 1.9.

- a. Explain why PCs and HPCs were evolutionary rather than revolutionary in the past 30 years**

Компјутерите и HPC-те еволуираа постепено во текот на изминатите 30 години поради континуираните подобрувања во веќе постојните технологии, а не поради радикалните, револуционерни промени.
- b. Discuss the drawbacks in disruptive changes in processor architecture. Why is the memory wall a major problem in achieving scalable changes in performance?**

Нарушувачките промени во архитектурата на процесорот може да доведат до проблеми со компатибилноста и да бараат значително софтверско реинженерство. Меморискиот ѕид е предизвик бидејќи процесорите ја надминаа брзината на меморијата, што доведува до побавни перформанси.
- c. Explain why x-86 processors are still dominating the PC and HPC markets.**

x86 процесорите продолжуваат да доминираат на пазарите за компјутери и HPC поради нивната широко распространета компатибилност со постоечкиот софтвер и инфраструктура, како и континуираниот напредок во перформансите и ефикасноста.

#### 1.10.

- a. What are the differences between multicore CPUs and GPUs in terms of architecture and usage?**

Повеќејадрените процесори имаат неколку моќни јадра оптимизирани за секвенцијална обработка, додека графичките процесори имаат многу поедноставни јадра дизајнирани за паралелни задачи како графичко прикажување и научни пресметки.

- b. Explain why parallel programming cannot match the progress of processor technology.

Паралелното програмирање се соочува со предизвици на ефикасно користење на повеќе јадра поради сложеноста во синхронизирањето на задачите и управувањето со споделените ресурси.

- c. Suggest ideas and defend your argument with some plausible solutions to this mismatch problem between core scaling and effective programming and use of multicores.

Едно решение е унапредување на алатките за автоматска паралелизација и програмски модели кои ја објаснуваат сложеноста, што им олеснува на програмерите ефективно да ги користат повеќејадрените архитектури.

- d. Explain why flash memory SSD can deliver better speedups in some HPC or HTC applications.

Флеш мемориските SSD се одлични во апликациите од HPC или HTC поради нивното мало доцнење и високата пропусна моќ, што значително го намалува времето за пронаоѓање податоци и ја забрзува обработката.

- e. Justify the prediction that InfiniBand and Ethernet will continue to dominate the HPC market.

Се предвидува дека InfiniBand и Ethernet ќе продолжат да доминираат на пазарот на HPC поради нивните можности за меѓусебно поврзување со голема брзина, воспоставените екосистеми и тековниот технолошки напредок.

#### 1.11.

Processor Micro-architectures	Architecture Characteristics	Advantages/Shortcomings	Representative Processors
<b>Single-threaded Superscalar</b>	<ul style="list-style-type: none"> <li>- Една инструкција по clock cycle.</li> <li>- Искористува паралелизам на ниво на инструкција во нишка.</li> <li>- Повеќекратните единици овозможуваат повеќе инструкции.</li> </ul>	<p><b>Предности:</b></p> <ul style="list-style-type: none"> <li>- Високи перформанси за една нишка</li> <li>- Ефикасни за задачи со низок паралелизам.</li> </ul> <p><b>Недостатоци:</b></p> <ul style="list-style-type: none"> <li>- Ограничена приспособливост во однос на перформансите.</li> </ul>	<ul style="list-style-type: none"> <li>- Intel Pentium 4 (NetBurst architecture)</li> <li>- AMD Athlon XP (K7 architecture)</li> </ul>
<b>Fine-grain Multithreading</b>	<ul style="list-style-type: none"> <li>- Наизменично помеѓу повеќе нишки.</li> <li>- Секоја нишка добива дел од процесорот.</li> <li>- Ги менува нишките на fine-grain интервали.</li> </ul>	<p><b>Предности:</b></p> <ul style="list-style-type: none"> <li>- Подобрено искористување на ресурсите.</li> <li>- Добро за оптоварување со повеќе задачи.</li> </ul> <p><b>Недостатоци:</b></p> <ul style="list-style-type: none"> <li>-Зголемена сложеност за</li> </ul>	<ul style="list-style-type: none"> <li>- Intel Hyper-Threading Technology (Pentium 4 HT)</li> <li>- Sun UltraSPARC T1 ("Niagara" processor)</li> </ul>

		управување со нишки.	
<b>Coarse-grain Multithreading</b>	<ul style="list-style-type: none"> <li>- Се префрла помеѓу нишки во погуби интервали.</li> <li>-Обезбедува повеќе ресурси за секоја нишка.</li> <li>-Се фокусира на извршување на поголем дел од инструкции.</li> </ul>	<p><b>Предности:</b></p> <ul style="list-style-type: none"> <li>- Ефикасно за апликации со подолги фази.</li> <li>-Ги намалува трошоците за thread switching.</li> </ul> <p><b>Недостатоци:</b></p> <ul style="list-style-type: none"> <li>-Може недоволно да ги искористи ресурсите за кратки задачи.</li> </ul>	<ul style="list-style-type: none"> <li>-IBM Power4 (Dual-core POWER4 processor)</li> <li>- Intel Xeon Phi (Knights Landing architecture)</li> </ul>
<b>Simultaneous Multithreading (SMT)</b>	<ul style="list-style-type: none"> <li>-Повеќе нишки истовремено.</li> <li>-Динамично преплетување на инструкции од различни нишки.</li> <li>-Ефикасно ги дели ресурсите помеѓу нишките.</li> </ul>	<p><b>Предности:</b></p> <ul style="list-style-type: none"> <li>- Искористеност на ресурси и пропусната моќ.</li> <li>- Ефикасно за оптоварувања со висок паралелизам на ниво на нишки.</li> </ul> <p><b>Недостатоци:</b></p> <ul style="list-style-type: none"> <li>- SMT може да предизвика конфликт за споделени ресурси.</li> </ul>	<ul style="list-style-type: none"> <li>-Intel Core i7 (Nehalem microarchitecture)</li> <li>-AMD Ryzen (Zen microarchitecture)</li> </ul>
<b>Multicore Chip Multiprocessor (CMP)</b>	<ul style="list-style-type: none"> <li>-Повеќе независни јадра на еден чип.</li> <li>-Секое јадро може самостојно да изврши нишка.</li> <li>-Идеален за паралелно извршување на повеќе задачи.</li> </ul>	<p><b>Предности:</b></p> <ul style="list-style-type: none"> <li>-Скалабилни перформанси преку паралелизам.</li> <li>-Ефикасно за мултитаскинг и апликации со повеќе нишки.</li> </ul> <p><b>Недостатоци:</b></p> <ul style="list-style-type: none"> <li>-Зголемена сложеност во софтверот за целосно искористување на повеќе јадра.</li> </ul>	<ul style="list-style-type: none"> <li>-Intel Core i7 (Multi-core variants, e.g., Core i7-8700K)</li> <li>-AMD Ryzen (Multi-core variants, e.g., Ryzen 7 3800X)</li> </ul>

## 1.12.

### a. Why are virtual machines and virtual clusters suggested in cloud computing systems?

Виртуелните машини и виртуелните кластери се препорачуваат во cloud системите бидејќи овозможуваат ефикасно искористување на ресурсите, изолација помеѓу оптоварувањата и флексибилност во зголемувањето или намалувањето на ресурсите врз основа на побарувачката.

**b. What breakthroughs are required to build virtualized cloud systems cost-effectively?**

Потребните откритија за исплатливи виртуелизирани cloud системи вклучуваат напредок во технологијата за виртуелизација за подобрени перформанси и ефикасност на ресурсите, како и иновации во хардверската и софтверската оптимизација.

**c. What are the impacts of cloud platforms on the future of the HPC and HTC industry?**

Cloud платформите ја трансформираат индустријата HPC и HTC преку изменување на пристапот до компјутерски ресурси со високи перформанси, овозможувајќи им на помалите организации да користат моќни машини.

**1.13.**

**a. What is an IaaS (Infrastructure-as-a-Service) cloud? Give one example system.**

Infrastructure-as-a-Service (IaaS) обезбедува виртуелизирани компјутерски ресурси преку интернет, овозможувајќи им на корисниците да изнајмуваат сервери, меморија и мрежни компоненти. Пример систем се веб-услугите на Amazon (AWS).

**b. What is a PaaS (Platform-as-a-Service) cloud? Give one example system.**

Platform-as-a-Service (PaaS) нуди платформа и средина за програмерите да градат и оджуваат апликации без потреба од управување со основната инфраструктура. Пример систем е Google Cloud Platform (GCP).

**c. What is a SaaS (Software-as-a-Service) cloud? Give one example system.**

Software-as-a-Service (SaaS) испорачува софтверски апликации преку интернет, без потреба од инсталација или одржување. Пример систем е Salesforce, платформа за управување со клиенти (CRM).

**1.17.**

Графичките процесори (GPUs) се извонредни во паралелната обработка, што ги прави ефикасни за задачи како што се графичко прикажување и тренирање на модели за машинско учење. Тие нудат огромни перформанси, но можат да бидат помалку програмибилни од процесорите. Процесорите се разноврсни и силни во последователните задачи. Во однос на MPP архитектурите, изборот на вистинскиот процесор зависи од специфичните цели за изведба, барањата за ефикасност и ограничувањата на апликацијата. Различни архитектури на MPP, како што се SIMD и MIMD, имаат различни јаки страни и компромиси во однос на стилот на обработка и приспособливоста.

**1.18.**

Атомеба е ОС базиран на микрокernels познат по своите дистрибуирани способности, без комерцијален успех. DCE, иако широко се користи за интероперабилност, се соочи со предизвици во приспособливоста кон технологиите што се развиваат. MOSIX е систем базиран на Линукс кој обезбедува миграција на процесите за балансирано оптоварување, кој се користи во апликации во облак средини. Сепак, се соочува со конкуренција од други технологии за виртуелизација. Сите три системи имаат



ограничено комерцијално усвојување поради различни причини, како што се сложеноста, проблемите со компатибилноста и технолошкиот развојот. Нивното влијание е поизразено во експериментални и истражувачки средини наместо во комерцијални апликации.

**Референци:**

- Distributed System Models and Enabling Technologies - CHAPTER 1
- <https://www.techtarget.com/searchcloudcomputing/definition/cloud-computing>
- <https://www.geeksforgeeks.org/characteristics-of-cloud-computing/>
- <https://www.embedded.com/how-to-calculate-cpu-utilization/>
- <https://www.tutorialspoint.com/difference-between-cluster-computing-and-grid-computing#:~:text=Nodes%20or%20computers%20can%20be,have%20homogeneous%20or%20heterogeneous%20network.&text=Computers%20of%20Cluster%20Computing%20are,resources%20to%20do%20other%20tasks.>
- <https://www.netapp.com/data-storage/high-performance-computing/what-is-hpc/>
- <https://www.micron.com/products/advanced-solutions/advanced-computing-solutions/hpc-desktop-systems>
- <https://www.acecloudhosting.com/blog/gpu-vs-cpu-high-performance-computing/>
- <https://www.nvidia.pl/content/cudazone/cudau/courses/ucdavis/lectures/tlp2.pdf>
- [https://en.wikipedia.org/wiki/Simultaneous\\_multithreading](https://en.wikipedia.org/wiki/Simultaneous_multithreading)
- <https://www.bmc.com/blogs/saas-vs-paas-vs-iaas-whats-the-difference-and-how-to-choose/>
- <https://www.quora.com/How-do-you-compare-three-distributed-operating-systems-Amoeba-DCE-and-MOSIX>
- <https://www.numerade.com/ask/question/compare-three-distributed-operating-systems-amoeba-dce-and-mosix-research-their-recent-developments-and-their-impact-on-applications-in-clusters-grids-and-clouds-discuss-the-suitability-of-e-98112/>