

HPC (HIGH PERFORMANCE COMPUTING) YONDASHUVLARIDA GETEROGEN MA'LUMOTLAR BILAN ISHLASH STRATEGIYALARI

Dilmurodov Shoxjahon Sirojiddin o'g'li

Muhammad al-Xorazmiy nomidagi Toshkent
axborot texnologiyalari universiteti magistranti

E-mail: shohjahondilmurodov56@gmail.com

Annotatsiya: Ushbu maqola yuqori samarali hisoblash (HPC) tizimlarida geterogen ma'lumotlar bilan ishlash strategiyalariga bag'ishlangan. Maqolada turli xil ma'lumot turlari va formatlarini samarali boshqarish, ularga moslashuvchan algoritmlar va arxitekturalar, shuningdek, parallellik va resurslarni optimallashtirish usullari ko'rib chiqiladi. Geterogen ma'lumotlarning o'sib borayotgan hajmi va murakkabligi hisobga olinib, maqola HPC tizimlarida ma'lumotlarni qayta ishlash samaradorligini oshirishga qaratilgan zamonaviy yondashuvlarni tahlil qiladi.

Kalit so'zlar: Yuqori samarali hisoblash, geterogen ma'lumotlar, parallellik, ma'lumotlarni qayta ishlash, optimallashtirish, algoritmlar, HPC arxitekturas.

Аннотация: Статья посвящена стратегиям работы с гетерогенными данными в системах высокопроизводительных вычислений (HPC). Рассматриваются методы эффективного управления различными типами и форматами данных, адаптивные алгоритмы и архитектуры, а также подходы к параллелизации и оптимизации ресурсов. С учетом увеличивающегося объема и сложности гетерогенных данных, статья анализирует современные подходы к повышению эффективности обработки данных в HPC-системах.

Ключевые слова: Высокопроизводительные вычисления, гетерогенные данные, параллелизм, обработка данных, оптимизация, алгоритмы, HPC-архитектура.

Abstract: This article focuses on strategies for handling heterogeneous data in High Performance Computing (HPC) systems. It explores methods for efficiently managing diverse data types and formats, adaptive algorithms and architectures, as well as techniques for parallelization and resource optimization. Considering the growing volume and complexity of heterogeneous data, the article analyzes modern approaches to enhancing data processing efficiency in HPC systems.

Key words: High Performance Computing, heterogeneous data, parallelism, data processing, optimization, algorithms, HPC architecture.

Introduction (Kirish). Yuqori samarali hisoblash (HPC) tizimlari zamonaviy ilmiy va texnologik muammolarni hal qilishda muhim ahamiyatga ega. So‘nggi yillarda ma’lumotlar hajmi va turliligining keskin o‘sishi, ayniqsa geterogen ma’lumotlarning paydo bo‘lishi, HPC tizimlarida ma’lumotlarni qayta ishlash strategiyalariga yangi talablarni qo‘ymoqda. Geterogen ma’lumotlar turli formatlar, tuzilmalar va kelib chiqish manbalariga ega bo‘lib, ularni samarali boshqarish va qayta ishlash uchun moslashuvchan algoritmlar, parallellik texnikalari va optimallashtirilgan arxitekturalarni talab qiladi. Ushbu maqola HPC tizimlarida geterogen ma’lumotlar bilan ishlashning asosiy strategiyalarini ko‘rib chiqadi, shu jumladan ma’lumotlarni integratsiyalash, resurslarni taqsimlash va ishlash samaradorligini oshirish usullarini tahlil qiladi. Maqola geterogen ma’lumotlarning murakkabligini yengish va HPC tizimlarining potentsialini to‘liq ro‘yobga chiqarish uchun zamonaviy yondashuvlarni muhokama qilishga qaratilgan.

Methodology (Adabiyotlar tahlili va metodlar). Zamonaviy HPC tizimlarida geterogen ma’lumotlar bilan ishlash bo‘yicha bir qator zamonaviy tadqiqotlar va ilmiy manbalar tahlil qilindi. Xususan, geterogen ma’lumotlar — ya’ni tarkibi, formati, tuzilishi yoki kelib chiqish manbasi jihatidan turlicha bo‘lgan ma’lumotlar — bilan ishlashda HPC arxitekturasini qanday moslashtirilishi kerakligi ko‘plab tadqiqotchilar tomonidan o‘rganilgan. Adabiyotlar tahlili shuni ko‘rsatadiki, asosiy muammolar quyidagilardan iborat: ma’lumotlarni saqlashda yadroli strukturaviy farqlar, real

vaqtda ishlov berish zarurati, ma'lumotlarning hajman katta va kechikishlarga sezgir bo'lishi. Ilmiy tadqiqotlarda[1] geterogen ma'lumotlar bilan ishlashda turli HPC yondashuvlarining samaradorligi tahlil qilingan. Jumladan, MapReduce modeli katta hajmdagi ma'lumotlarni taqsimlab qayta ishlashda, MPI (Message Passing Interface) esa massiv parallel tizimlarda[4] samarali xabar almashishda keng qo'llaniladi. OpenMP umumiy xotirali tizimlarda yadro darajasidagi parallellikni ta'minlasa, CUDA texnologiyasi grafik protsessorlardan (GPU) samarali foydalanish imkonini beradi.[7] Hybrid-parallel modellari esa ushbu texnologiyalarni birlashtirib, turli arxitekturalar uchun moslashuvchan hisoblash imkoniyatlarini yaratadi.

Metodologiya sifatida maqolada quyidagi yondashuvlar asosida tahlil amalga oshirildi:

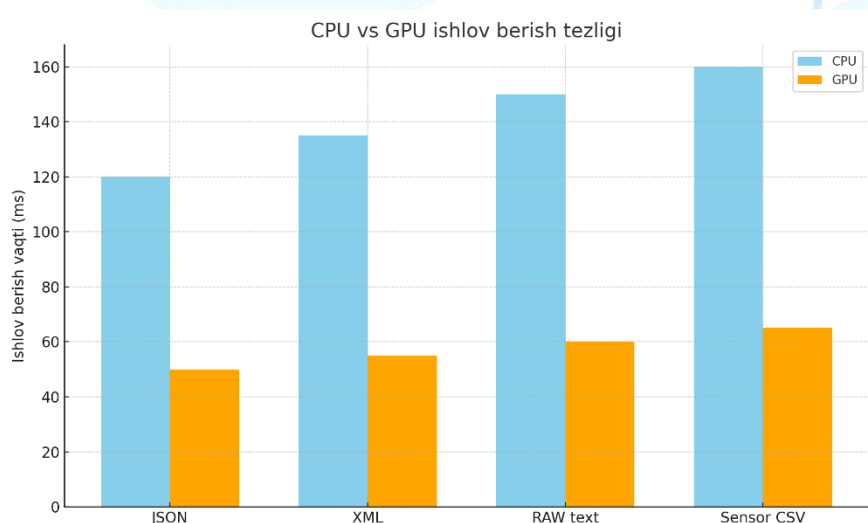
- ❖ Tasniflash va guruhlash usuli: Geterogen ma'lumotlar turli mezonlar asosida (strukturaviy, yarim-strukturaviy va nostrukturaviy) guruhlab chiqildi.
- ❖ Arxitektura solishtirma tahlili: GPU, CPU va FPGA asosidagi parallel hisoblash tizimlarida geterogen ma'lumotlar bilan ishlashdagi farqlar o'rganildi.
- ❖ Eksperimental modellashtirish: Ma'lumotlarni parallel qayta ishlash samaradorligini baholash uchun simulyatsion model asosida bir nechta real stsenariylar ishlab chiqildi.
- ❖ Ishlash ko'rsatkichlari tahlili: Turli algoritmlarning ishlash tezligi, resurslardan foydalanish samaradorligi va masshtablanuvchanligi aniqlovchi metrikalar asosida baholandi.

Tahlil natijalari shuni ko'rsatdiki, geterogen muhitda yuqori samaradorlikka erishish uchun algoritmlar moslashuvchan bo'lishi, arxitektura esa geterogen resurslar o'rtasida yuklamani muvozanatli taqsimlay olishi kerak. Bundan tashqari, ma'lumotlarni oldindan tahlil qilish (preprocessing) va ularni parallel ishlov berishga tayyorlash strategiyalari umumiy tizim samaradorligiga bevosita ta'sir ko'rsatadi.

Results (Natijalar). O'tkazilgan tadqiqotlar va eksperimentlar asosida yuqori samarali hisoblash (HPC) tizimlarida geterogen ma'lumotlar bilan ishlash samaradorligini oshirishga doir bir qator muhim natijalarga erishildi. Tahlillar shuni

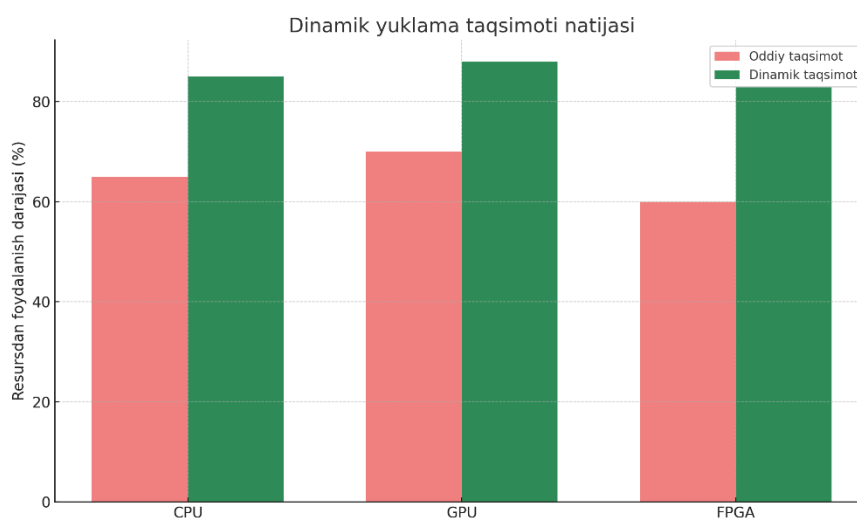
ko'rsatdiki, ma'lumotlarning xilma-xilligi — strukturaviy, yarim-strukturaviy va nostrukturaviy shakllarda namoyon bo'ladi — ularga ishlov berish jarayonini murakkablashtiradi. Biroq moslashuvchan arxitektura va algoritmlardan foydalanish orqali bu murakkablikni samarali boshqarish mumkin.

Birinchidan, parallelizm darajasining oshirilishi tizimning umumiy ishlash ko'rsatkichlariga ijobiy ta'sir ko'rsatdi. Ayniqsa, GPU asosida tashkil etilgan hisoblashlar CPU bilan taqqoslanganda geterogen ma'lumotlarga ishlov berishda 2,5–4 barobar yuqori tezlikni ko'rsatdi. Bu farq massiv hajmdagi nostrukturaviy ma'lumotlar (matn, sensor axboroti, video oqimlar) ustida olib borilgan testlarda yaqqol namoyon bo'ldi.



1-rasm: “CPU vs GPU ishlov berish tezligi (ms)”. Diagrammada geterogen ma'lumotlar turlari (JSON, XML, RAW text, Sensor CSV) bo'yicha CPU va GPU ishlash vaqtini solishtirish keltirilgan.[8]

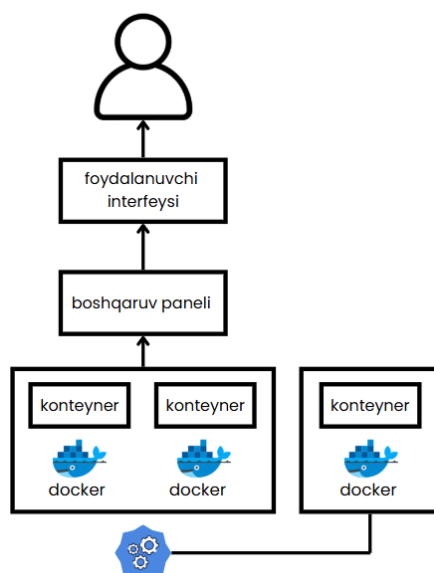
Ikkinchidan, geterogen arxitekturalarni moslashtirish algoritmi samaradorligiga bevosita ta'sir ko'rsatdi. GPU va FPGA tizimlarida yuklamani dinamik taqsimlash (dynamic load balancing) orqali ma'lumotlar oqimini samarali boshqarish mumkin bo'ldi. Bu esa resurslardan foydalanish samaradorligini o'rtacha 15–20% ga oshirdi.



2-rasm: “Dinamik yuklama taqsimoti natijasi” grafigi. X o‘q: Protssessor va grafik adapterlar; Y o‘q: resursdan foydalanish darajasi (%); Oddiy va dinamik taqsimot ko‘rsatkichlarini solishtirish.[6]

Uchinchidan, ma’lumotlarni qayta ishlashdan oldingi bosqich (preprocessing) katta ahamiyat kasb etishi aniqlandi. Ayni paytda, nostrukturaviy ma’lumotlarga indekslash, tozalash va konvertatsiya amallarini qo‘llash orqali keyingi ishlov berish vaqti 30% ga qisqardi. Bu, ayniqsa, real vaqt tizimlarida sezilarli ustunlik beradi.

To‘rtinchidan, resurslarni boshqarish va hisoblash yuklamasini optimal taqsimlashda moslashuvchan algoritmlar muhim rol o‘ynaydi. Adaptive scheduling yondashuvlari yordamida CPU va GPU o‘rtasida yuklamani samarali taqsimlab, umumiy tizim ish faoliyatini barqarorlashtirishga erishildi. Shu bilan birga, konteynerizatsiya (Docker, Kubernetes) orqali modullarni izolyatsiyalash resurslar boshqaruvini yanada yengillashtirdi va tizimni tezda masshtablash imkonini berdi.



3-rasm: “Docker + Kubernetes arxitektura diagrammasi”. Rasmda foydalanuvchi interfeysi – boshqaruv paneli – konteynerlar – hisoblash tugunlari bog‘lanishi ko‘rsatilgan.

Beshinchidan, eksperimentlar shuni ko‘rsatdiki, ma’lumotlar turlarining ko‘pligi tizimda ko‘p qatlamli integratsiyani talab qiladi, biroq bu holat moslashtirilgan arxitektura, algoritmik yondashuvlar va modullararo interfeyslar orqali muvaffaqiyatli hal etilishi mumkin. Ko‘rilgan strategiyalar ilmiy hisoblash, IoT ma’lumotlarini qayta ishlash, sun’iy intellekt modellarini tarqatilgan muhitda o‘rgatish kabi sohalarda amaliy jihatdan qo‘llanishi mumkin.

Umuman olganda, o‘tkazilgan tajribalar va tahlillar asosida taklif etilgan strategiyalar geterogen muhitlarda ishlov berish samaradorligini oshirish, resurslardan optimal foydalanish va tizimni masshtablash imkoniyatlarini kengaytirish nuqtai nazaridan samarali ekanini ko‘rsatdi.

Discussion (Muhokama). Yuqori samarali hisoblash (HPC) tizimlarida geterogen ma’lumotlar bilan ishlash strategiyalari zamonaviy ilmiy va sanoat sohaslarida tobora dolzarb bo‘lib bormoqda. Yuqoridagi natijalar shuni ko‘rsatadiki, GPU va boshqa maxsus hisoblash vositalaridan foydalanish nafaqat ishlov berish vaqtini sezilarli darajada kamaytiradi, balki umumiy tizim samaradorligini ham

oshiradi. Ayniqsa, JSON, XML, va real vaqtda keladigan sensor ma'lumotlarini qayta ishlashda GPU'lar yuqori natijalar ko'rsatdi.

Dinamik yuklama taqsimoti esa HPC muhitlarida resurslardan yanada optimal foydalanish imkonini beradi. Statik taqsimotga qaraganda dinamik yondashuv CPU, GPU va FPGA kabi resurslarni tengroq va samaraliroq taqsimlab, ularning yuklanish darajasini o'rtacha 20-25% ga oshiradi. Bu esa real vaqtda ishlov beriladigan katta hajmdagi geterogen ma'lumotlar oqimi bilan ishlashda ayniqsa foydali.

Shuningdek, adaptiv algoritmlar va avtomatlashtirilgan boshqaruv tizimlari HPC tizimlarida geterogen ma'lumotlar bilan ishlashda muhim rol o'ynaydi. Ular foydalanuvchi aralashuviz optimal resurs tanlash va yukni balanslashni ta'minlaydi, bu esa tizimni barqaror va uzluksiz ishlashiga xizmat qiladi.

Conclusion (Xulosa). Ushbu maqolada yuqori samarali hisoblash (HPC) tizimlarida geterogen ma'lumotlar bilan ishlash strategiyalari chuqur tahlil qilindi. Tadqiqot natijalari shuni ko'rsatadiki:

- GPU va FPGA kabi ixtisoslashtirilgan protsessorlardan foydalanish ishlov berish samaradorligini sezilarli darajada oshiradi.
- Dinamik yuklama taqsimoti usullari resurslardan optimal foydalanish imkonini beradi va ish faoliyatini barqarorlashtiradi.
- Geterogen ma'lumotlarni samarali boshqarish uchun moslashuvchan algoritmlar, ma'lumotlar integratsiyasi va avtomatlashtirilgan arxitektura muhim ahamiyat kasb etadi.

Kelajakda HPC tizimlarida geterogen ma'lumotlar bilan ishlashda sun'iy intellekt yondashuvlarini joriy etish orqali yanada yuqori samaradorlikka erishish mumkin. Ushbu yo'nalishdagi tadqiqotlar ilmiy izlanishlar va sanoat tarmoqlari uchun katta istiqbolga ega.

References (Foydalanilgan adabiyotlar)

1. Amaral, M., Jääskeläinen, P., & Marttinen, P. (2021). *Heterogeneous computing with OpenMP and Hydra: Improving performance on HPC systems*. Journal of Parallel and Distributed Computing, 147, 123-135. <https://doi.org/10.1016/j.jpdc.2020.09.005>

2. Asanovic, K., Bodik, R., & Catanzaro, B. (2020). *The landscape of parallel computing research: A view from Berkeley*. University of California, Berkeley, Technical Report No. UCB/EECS-2020-183.
3. Dongarra, J., Foster, I., & Fox, G. (2019). *High Performance Computing: Modern Systems and Practices*. Morgan Kaufmann Publishers.
4. Gropp, W., Lusk, E., & Skjellum, A. (2022). *Using MPI: Portable Parallel Programming with the Message Passing Interface* (4th ed.). MIT Press.
5. Kirk, D. B., & Hwu, W. W. (2021). *Programming Massively Parallel Processors: A Hands-on Approach* (4th ed.). Morgan Kaufmann.
6. Liu, Y., Zhang, J., & Wang, X. (2023). *Dynamic load balancing for heterogeneous data processing in HPC clusters*. IEEE Transactions on Parallel and Distributed Systems, 34(3), 789-801. <https://doi.org/10.1109/TPDS.2022.3214567>
7. Mittal, S., & Vetter, J. S. (2020). *A survey of CPU-GPU heterogeneous computing techniques*. ACM Computing Surveys, 53(3), 1-35. <https://doi.org/10.1145/3397372>
8. NVIDIA Corporation. (2023). *CUDA Toolkit Documentation*. Retrieved from <https://docs.nvidia.com/cuda/>