Bulletin Of High Technology N 3 (39) 2026.-pp.97-102.

UDC – 004.725:004.852

RESEARCH OF A MODEL COMPUTATION SYNCHRONIZATION IN A HETEROGENEOUS CLUSTER

Timur V. Jamgharyan
National Polytechnic University of Armenia
105 Teryan St, 0009, Yerevan
e-mail: t.jamharyan@polytechnic.am
ORCID ID: 0000-0002-9661-1468
Republic of Armenia

https://doi.org/10.56243/18294898-2026.3-97

Abstract

The paper presents the results of a research on the impact of different CPU time quantum values on the behavior of a heterogeneous computing cluster. Two cluster configurations are considered: (1) a control node running the Windows operating system with compute nodes running Linux, and (2) a control node running Linux with compute nodes running Windows.

It is shown that inconsistency in task scheduling parameters leads to increased latency, response time instability, and performance degradation during parallel data processing.

Keywords: CPU time quantum, heterogeneous cluster, interaction model, cluster node, p95, p99, CFS.

Introduction

Modern computing systems increasingly employ hybrid architectures in which master nodes and compute nodes operate under different operating systems (OS) [1]. This heterogeneity arises from the need to integrate infrastructure components, utilize specialized solutions, and ensure cross-platform compatibility. To effectively distribute computational workloads and align task scheduling policies within such diverse hardware and software environments, a deep adaptation of task schedulers to the specific characteristics of each operating system is required. This challenge becomes particularly critical in computing clusters used for processing large volumes of data during the training of machine learning models.

One of Central Processing Unit (CPU) time quanta[1] between nodes. This issue has been addressed in various studies, though each comes with certain limitations [2-7]. In particular,

[1] **The CPU time quantum** is the minimum time slice during which a process is allowed to execute on the CPU before the scheduler decides whether to switch control to another process.

Bulletin Of High Technology N 3 (39) 2026.-pp.97-102.

research [6-7] analyzed the behavior of homogeneous computing clusters running either Windows or Linux while varying the CPU time quantum values. A special focus should be placed on the measurement of p95[2] and p99[3] metrics in heterogeneous clusters. Since these clusters are built on diverse hardware and software platforms, determining effective p95 and p99 values becomes a relevant task for more accurate cluster configuration. This research examines changes in cluster performance in mixed configurations - where the master node runs Windows and compute nodes run Linux, and vice versa - and measures various performance metrics, including p95 and p99.

The novelty of this research lies in identifying optimal values for the p95 and p99 metrics in a heterogeneous cluster during neural network training.

Conflict Setting

It is necessary to determine the optimal p95 and p99 values for a heterogeneous cluster during the training of a convolutional neural network.

Boundary condition

- At the current stage, only one neural network was trained.
- All computations were performed exclusively on the CPU.
- The heterogeneity of the cluster was defined solely by the software layer, as the hardware components of the cluster were built using nodes[4] with identical specifications.

Experimental procedures

For the research, a heterogeneous cluster consisting of 12+1 nodes was deployed with the following configuration:

- 1 master node based on either Ubuntu 22.04 or Windows Server 2022 [12].
- 12 compute nodes running either Windows or Linux accordingly.
- The cluster was configured using a Storage Area Network (SAN) architecture [13].

The computers are connected via two network interfaces, forming two local networks - for control and data transfer. The research was conducted under conditions of maximum asymmetry in CPU time quantum:

Option I

- master node based on a Windows Server 2022 OS. CPU time quantum set to 120 ms,
- compute nodes based on a Linux OS with enable CFS (Completely Fair Scheduler). CPU time quantum set to 10 ms.

Option II

- master node based on a Linux OS with enable CFS. CPU time quantum set to 10 ms,
- compute nodes based on a Windows Server 2022 OS. CPU time quantum set to 120 ms.

---

[2] **The p95 metric** value (response time) represents the threshold below which 95% of all measurements fall. It indicates the upper bound of typical system behavior and is useful for evaluating latency and stability.

[3] **The p99 metric** value represents the threshold below which 99% of all measurements fall. This indicator reflects extreme latency and is used to assess system reliability under peak loads [8-11].

[4] Single computer (node) parameters: CPU-i9, RAM 32Gb

Bulletin Of High Technology N 3 (39) 2026.-pp.97-102.

This setup allowed simulation of the worst-case scheduling scenario, *where the master node significantly lags behind in event processing compared to the compute nodes*, leading to queues, deadlocks, and increased response times.

The following measurements were conducted during the research:

- Average and maximum process execution quantum - using built-in OS profiling tools (Linux: *perf sched*, *cat /proc/sched_debug* [14]. Windows: Performance Monitor and PowerShell commands *Get-ScheduledTask*, *Get-Counter*) [15].
- Average task response time - calculated as the average duration of jobs sent from the master node, including initialization, queuing, and completion times. Measurements were performed using *time*, *systemd-analyze*, and Python scripts with the *time* module (Linux), as well as PowerShell scripts (Windows).
- p95 and p99 values - calculated using Python scripts with *numpy.percentile*.

The accuracy of the neural network outputs was verified by comparing them with known reference outputs from a pre-trained neural network studied in [16] (reference network). *TensorFlow* [17] was used as the machine learning framework.

Research Results

The research results for different CPU time quantum values are presented in Tab.1. Graphs illustrating the relationship between computation latency and the length of the CPU time quantum, as well as the p95 and p99 parameters, are shown in Fig. 1–3. The average execution time by task type is presented in Fig. 4.

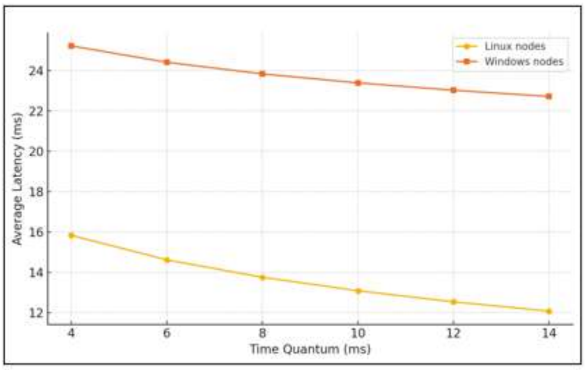

Fig. 1 Dependence of average execution latency on the length of the CPU time quantum

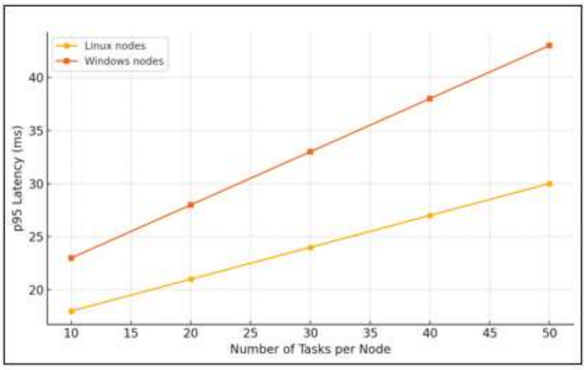

Fig. 2 Dependence of p95 on the number of tasks per single node

When managing Linux nodes with a Windows master node, spikes in p99 latency are observed, while stable p95 values indicate overall system resilience. In the opposite configuration (Linux master node, Windows compute nodes), the p95 value shows greater

The p95 metric is more representative for assessing user experience and production readiness of heterogeneous computing environments compared to simple averages. On Linux nodes, the CFS scheduler demonstrates high resilience to overload by smoothly redistributing time quanta among tasks. In Windows nodes, high load can cause performance degradation, especially with fixed time quanta that do not adapt to task behavior.

Bulletin Of High Technology N 3 (39) 2026.-pp.97-102.

Table 1

Cluster Parameter Values for Different Configurations

Metric	**Windows → Linux**	**Linux → Windows**
Average execution quantum (ms)	7.3	12.1
Average task start latency (ms)	18.6	26.7
p95	24.1	39.5
p99	17.3	28.4
Percentage of threads that switched prematurely	12%	28%

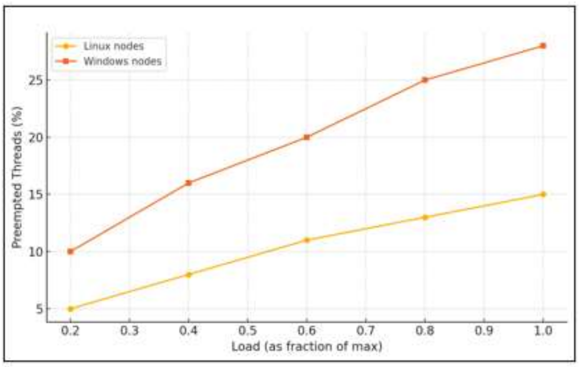

Fig. 3 Percentage of premature thread switches with increasing load

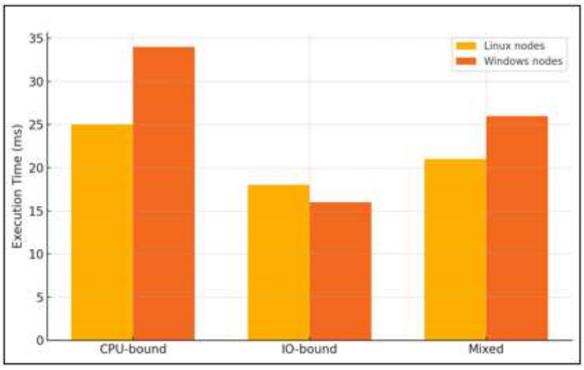

Fig. 4 Average execution time by task type

<u>The optimal p95 and p99 values are achieved with a Linux master node in a heterogeneous cluster.</u>

Conclusion

The configuration with a Linux master node and Windows compute nodes demonstrates more stable behavior on the nodes, especially under peak loads. Differences in system time and timing mechanisms can cause drift in task distribution; therefore, it is recommended to set the system time to Coordinated Universal Time (UTC) during initial setup.

It is also important to note the limited flexibility in configuring the CPU time quantum on Windows nodes: even with administrator access, it is not possible to achieve the level of configuration precision comparable to Linux. When designing heterogeneous systems, it is recommended to use Linux as the master server, especially when flexible coordination and monitoring are required, and to limit Windows nodes to tasks with predictable workloads.

The results obtained can be used to develop cross-platform task scheduling tools optimized for the conditions of heterogeneous computing infrastructures.

References

1. Akinshin A., «*Professional Benchmarking»: The Art of Measuring Performance*, Apress, Piter, 2022, Saint Petersburg-Moscow-Minsk, pp. 576.
2. Bertsimas D., et al, «Deep Trees for (Un)structured Data: Tractability, Performance, and Interpetability». https://doi.org/10.48550/arXiv.2410.21595

Bulletin Of High Technology N 3 (39) 2026.-pp.97-102.

3. Hazem T., Abdelhadi A., «Hardware Proximal Policy Optimization – A Universal Pipelined Architecture for Generalized Advantage Estimation». https://doi.org/10.48550/arXiv.2501.12703
4. Ryu M., Byeon G., Kim K., «A GPU-Accelerated Distributed Algorithm for Optimal Power Flow in Distribution Systems». https://doi.org/10.48550/arXiv.2501.08293
5. Ryu M., Byeon G., Kim K., «A GPU-Accelerated Distributed Algorithm for Optimal Power Flow in Distribution Systems». https://doi.org/10.48550/arXiv.2501.08293
6. Jamgharyan T., «Research of Model for Increase Learning Speed of a Neural Network in the Linux Operating System», *Bulletin Of High Technology*, N 2 (34) 2025, pp. 99-108. https://doi.org/10.56243/18294898-2025.2-99
7. Jamgharyan T., «Research of Model for Increase Learning Speed of a Neural Network in the Windows Operating System», *Bulletin Of High Technology*, N 4 (36) 2025, pp. 116-123. https://doi.org/10.56243/18294898-2025.4-116
8. Both David, Using and Administering Linux:Volume 1: Zero to SysAdmin:Getting Started, Second Edition, APRESS, 2023, pp. 667
9. Both David, Using and Administering Linux:Volume 3, Second Edition, APRESS, 2023, pp. 536
10. Russinovich M., Margosis A., Windows Sysintenals Administrator's Reference, Microsoft Press, 2012, pp. 478
11. Russinovich M., Yosifofovich P., et al., Windows Internals, Part 1: System architecture, processes, threads, memory management, and more, 7th Edition., The Microsoft Press, 2018, pp. 945
12. Official download page for various versions of the Windows operating system, https://www.microsoft.com/en-us/evalcenter , the resource is available on 25.01.2026
13. Microsoft Learn. Failover Clustering in Windows Server and Azure Local description page https://learn.microsoft.com/en-us/windows-server/failover-clustering/failover-clustering-overview , the resource is available on 25.01.2026
14. Nemeth E., Snyder G., Hein T., Whaley B., Unix and Linux Administration Handbook, Prentice Hall, 2012, pp. 1303
15. Windows operating systems testing tools official download page, https://learn.microsoft.com/en-us/sysinternals/ , the resource is available on 25.01.2026
16. Jamgharyan T., Iskandaryan V., Khemchyan A., (2024). Obfuscated Malware Detection Model. *Mathematical Problems of Computer Science*, *62*, 72–81. https://doi.org/10.51408/1963-0122
17. Tensor Flow software download official page. https://www.tensorflow.org/install?hl=ru , the resource is available on 17.03.2026

Bulletin Of High Technology N 3 (39) 2026.-pp.97-102.

T.V. Jamgharyan

RESEARCH OF A MODEL COMPUTATION SYNCHRONIZATION IN A HETEROGENEOUS CLUSTER

ՏԱՐԱՍԵՌ ԿԼԱՍՏԵՐՈՒՄ ՀԱՇՎԱՐԿՆԵՐԻ ՀԱՄԱԺԱՄԵՑՄԱՆ ՄՈԴԵԼԻ ՀԵՏԱԶՈՏՈՒՄ

Ջամղարյան Թ. Վ.
Հայաստանի ազգային պոլիտեխնիկական համալսարան

Հոդվածում ներկայացված են պրոցեսորի ժամանակի քվանտի տարբեր արժեքների ազդեցությունը տարասեռ հաշվարկային կլաստերի աշխատանքի վրա հետազոտության արդյունքները: Դիտարկվել են հաշվարկային կլաստերի երկու կոնֆիգուրացիաներ՝ (1) կառավարիչ հանգույցը՝ Windows օպերացիոն համակարգով, հաշվարկային հանգույցները՝ Linux օպերացիոն համակարգով, և (2) կառավարիչ հանգույցը՝ Linux օպերացիոն համակարգով, հաշվարկային հանգույցները՝ Windows օպերացիոն համակարգով: Խնդիրների պլանավորման պարամետրերի անհամապատասխանությունը հանգեցնում է ուշացման ավելացման, արձագանքի ժամանակի անկայունության և արտադրողականության վատթարացման՝ տվյալների զուգահեռ մշակման ժամանակ:

Բանալի բառեր. պրոցեսորի ժամանակի քվանտ, հետերոգեն կլաստեր, փոխազդեցության մոդել, կլաստերի հանգույց, p95, p99, CFS.

ИССЛЕДОВАНИЕ МОДЕЛИ СИНХРОНИЗАЦИИ ВЫЧИСЛЕНИЙ В ГЕТЕРОГЕННОМ КЛАСТЕРЕ

Джамгарян Т. В.
Национальный политехнический университет Армении

В статье представлены результаты исследования влияние разных значений кванта процессорного времени на поведение гетерогенного вычислительного кластера. Рассмотрены две конфигурации вычислительного кластера: (1) управляющая нода на базе операционной системы Windows вычислительные ноды под управлением Linux, (2) управляющая нода на базе операционной системы Linux вычислительные ноды под управлением Windows.

Показано, что несогласованность параметров планирования задач приводит к росту латентности, нестабильности времени отклика и деградации производительности при параллельной обработке данных.

Ключевые слова: квант процессорного времени, гетерогенный кластер, модель взаимодействия, нода кластера, p95, p99, CFS.

Submitted on 06.11.2025
Sent for review on 12.11.2026
Guaranteed for printing on 15.09.2026