High Performence Computing

7102CEN-3

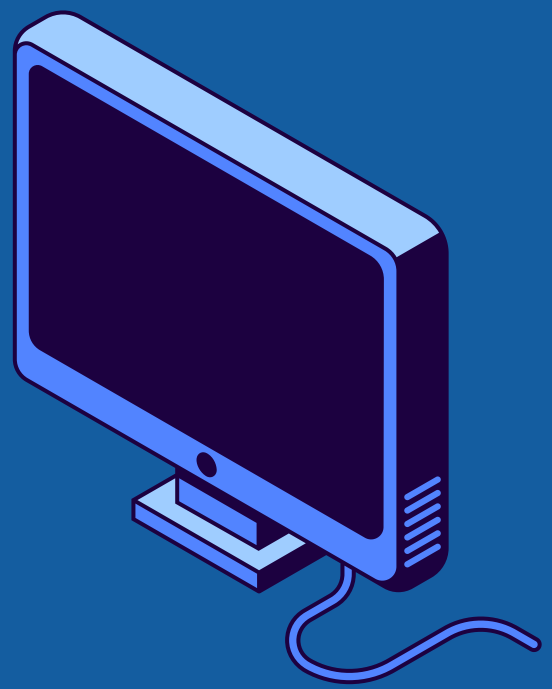
Sustainability and Energy Efficiency in HPC Data Centers

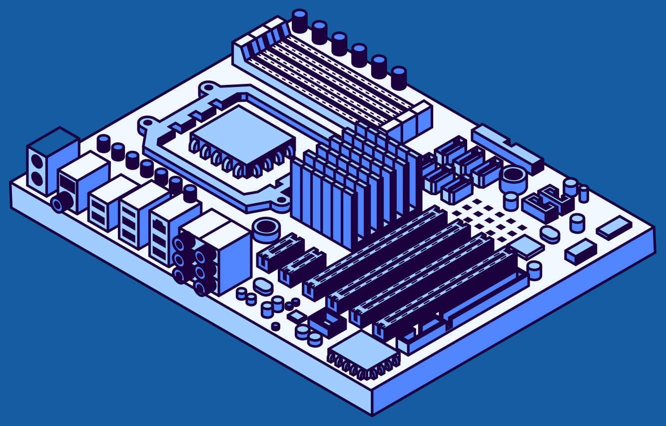
Andrea Chinnici 1, Eyvaz Ahmadzada 2,3, Ah-Lian Kor 2 , Davide De Chiara 4, Adrián Domínguez-Díaz 1 Luis de Marcos Ortega 1 andMartaChinnici3,

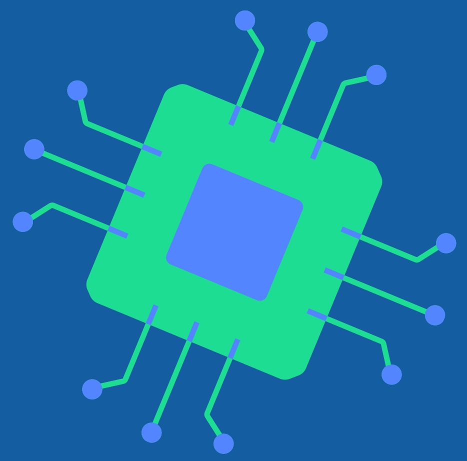
By: Hussain Hassan Alshikh

Supervised by: Prof. Paul Rodrigues

Outline







Introduction

HPC data centers are essential for AI, cloud computing, and big data but consume vast amounts of energy, leading to high costs and environmental impact. Efficient thermal management is crucial to optimize performance and reduce waste. This research focuses on CRESCO6, a high-performance computing (HPC) cluster at the European Nuclear Energy Agency (ENEA) data center in Italy, aiming to enhance energy efficiency using data analytics and machine learning. By predicting optimal cooling settings, the study improves sustainability, reduces overheating, and ensures smarter resource allocation for greener, cost-effective operations.

Introduction

CRESCO6 is one of Italy's largest HPC clusters, consisting of 434 nodes, each with Intel Xeon Platinum 8160 processors and 192GB RAM. It supports parallel computing, big data analysis, and AI-driven simulations, making it an ideal platform for energy efficiency research

Related Work

Historical Evolution of Data Centers:

- Transition from mainframes to cloud-based and AI-driven infrastructures.
- Software-Defined Networking (SDN) has improved efficiency.

Challenges in Energy Efficiency:

- High-density computing leads to thermal issues.
- Traditional cooling systems are inefficient.
- Idle nodes cause unnecessary energy waste.

Existing Strategies for Improvement:

- Liquid cooling, free-air cooling, and virtualization.
- Machine learning for workload balancing and predictive maintenance.

Methodology

Phase 1: Exploratory Data Analysis

- Data Collected:
  - Job schedules, cooling system data, sensor logs, and environmental conditions.
- Found that:
  - Job failures increase energy waste.
  - Seasonal variations impact cooling efficiency.

Phase 2: Inferential Statistical Analysis

- Key Findings:
  - Hot-aisle temperature correlates with fan speed and power consumption.
  - Higher temperatures lead to more energy use.

Methodology

Phase 3: Machine Learning for Cooling Optimization

- Models Used:
  - Ridge Regression to predict hot-aisle temperature.
  - AI-based fan speed and cooling control.
- Results:
  - Predicted temperature deviations within ±1.3°C.
  - improving efficiency.

Phase 4: Idle Node Shutdown and Thermal Impact

- Idle nodes contribute significantly to wasted energy.
- Simulation Results:
  - Shutting down idle nodes improves cooling efficiency.
  - Predicted vs. Actual temperature showed improved thermal stability.

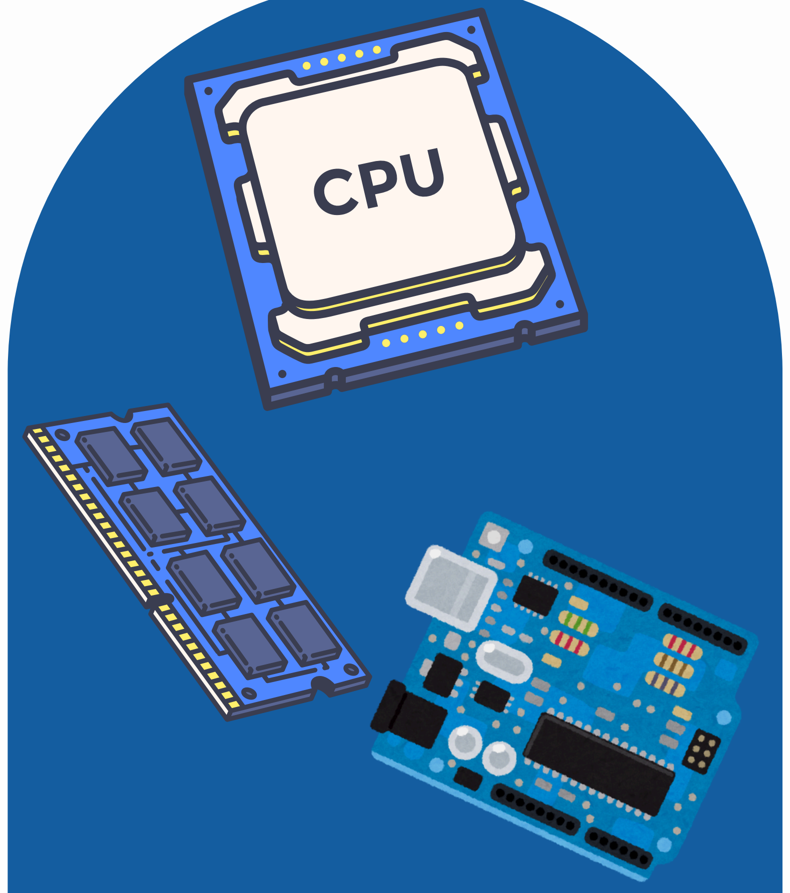
Results and Discussion

Cooling System Inefficiencies

- Detected inefficiencies in the early months of the study.
- AI-based optimizations effectively reduced energy consumption.

Data Insights

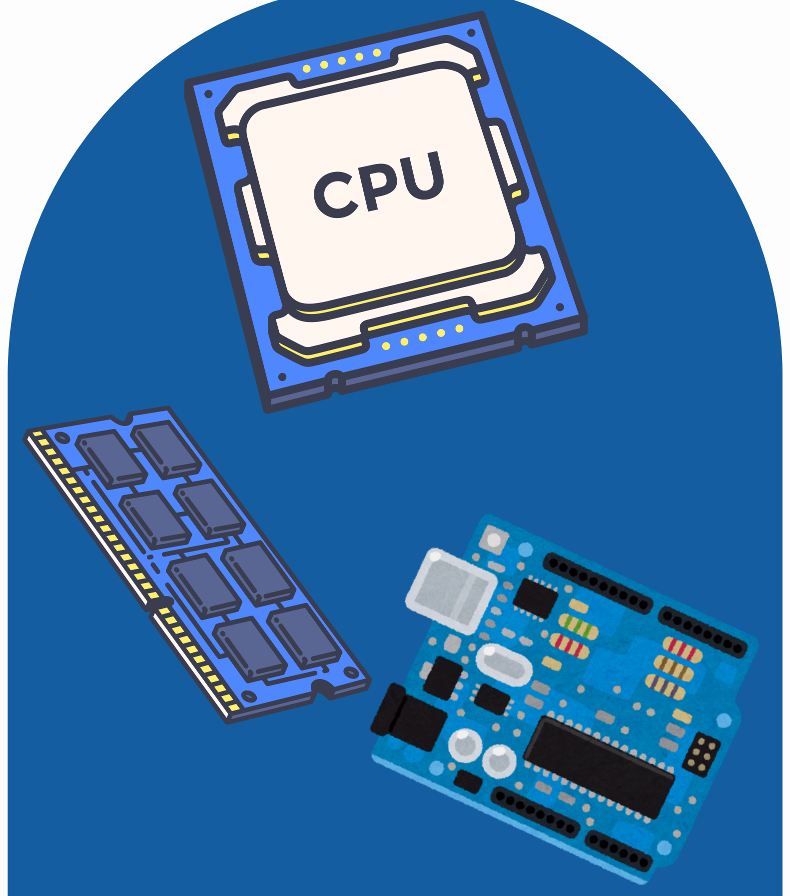
- Strong correlations:
  - Hot-aisle temp & return-air temp (0.94) → Higher heat buildup.
  - Hot-aisle temp & fan speed (-0.82) → Fan speed decrease suggests cooling inefficiencies.
- Impact: Machine learning should be used to optimize cooling settings.



Results and Discussion

Machine Learning Performance

- Comparison between actual vs. predicted hot-aisle temperatures.
- Model Accuracy:
  - MSE = 1.3156 → High prediction accuracy.
  - Suggests further improvements for real-time cooling adjustments.

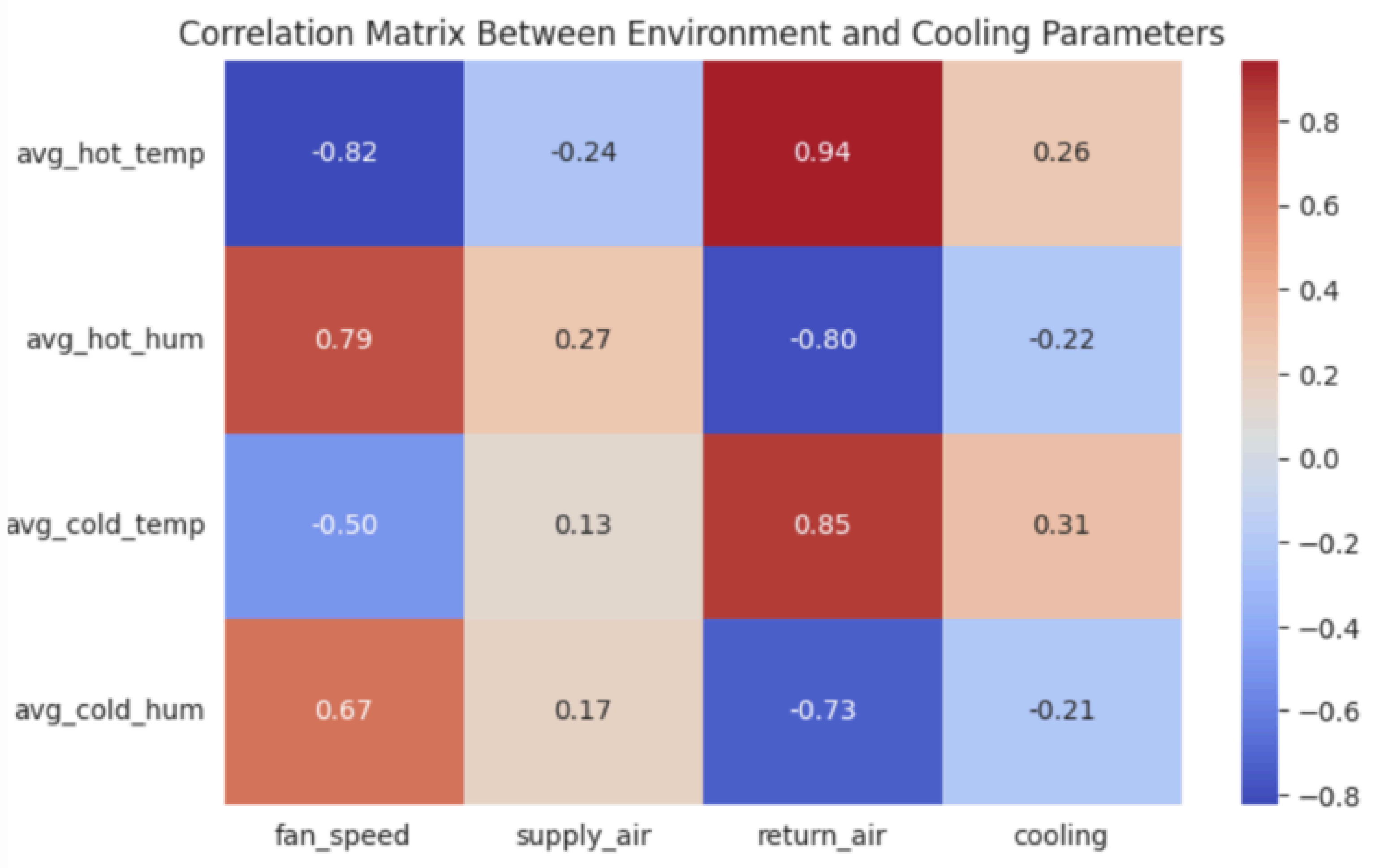


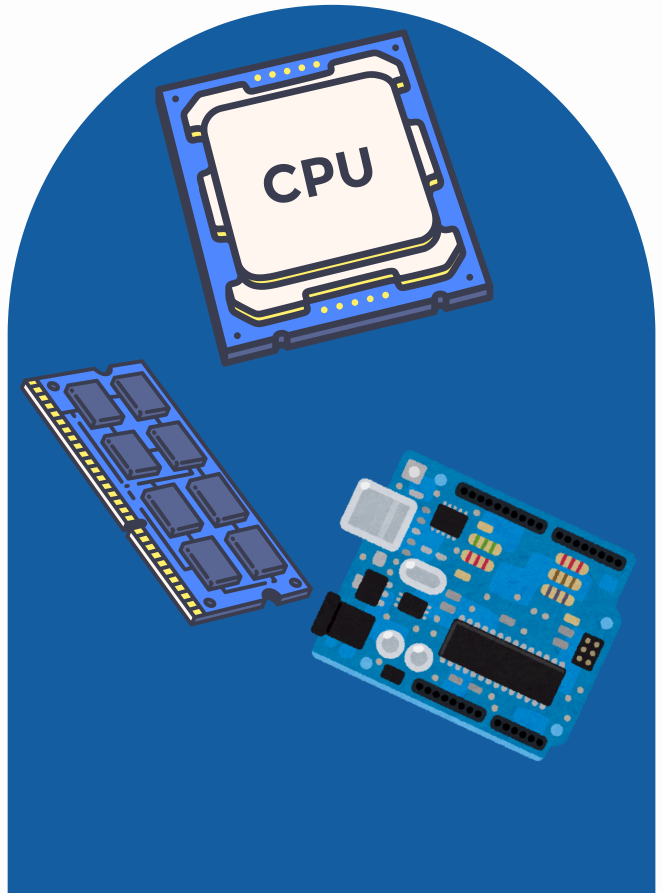
Results and Discussion

Purpose: Shows relationships between hot-aisle temperature, fan speed, and return-air temperature.

Key Insights:

- Hot-aisle & return-air temperature (0.94 correlation) → Higher heat buildup.
- Hot-aisle temp & fan speed (-0.82 correlation) → Fan speed decreases as heat rises, suggesting cooling inefficiencies.
- Impact: Cooling adjustments should be optimized using machine learning to improve energy efficiency and prevent overheating.





Results and Discussion

Purpose:

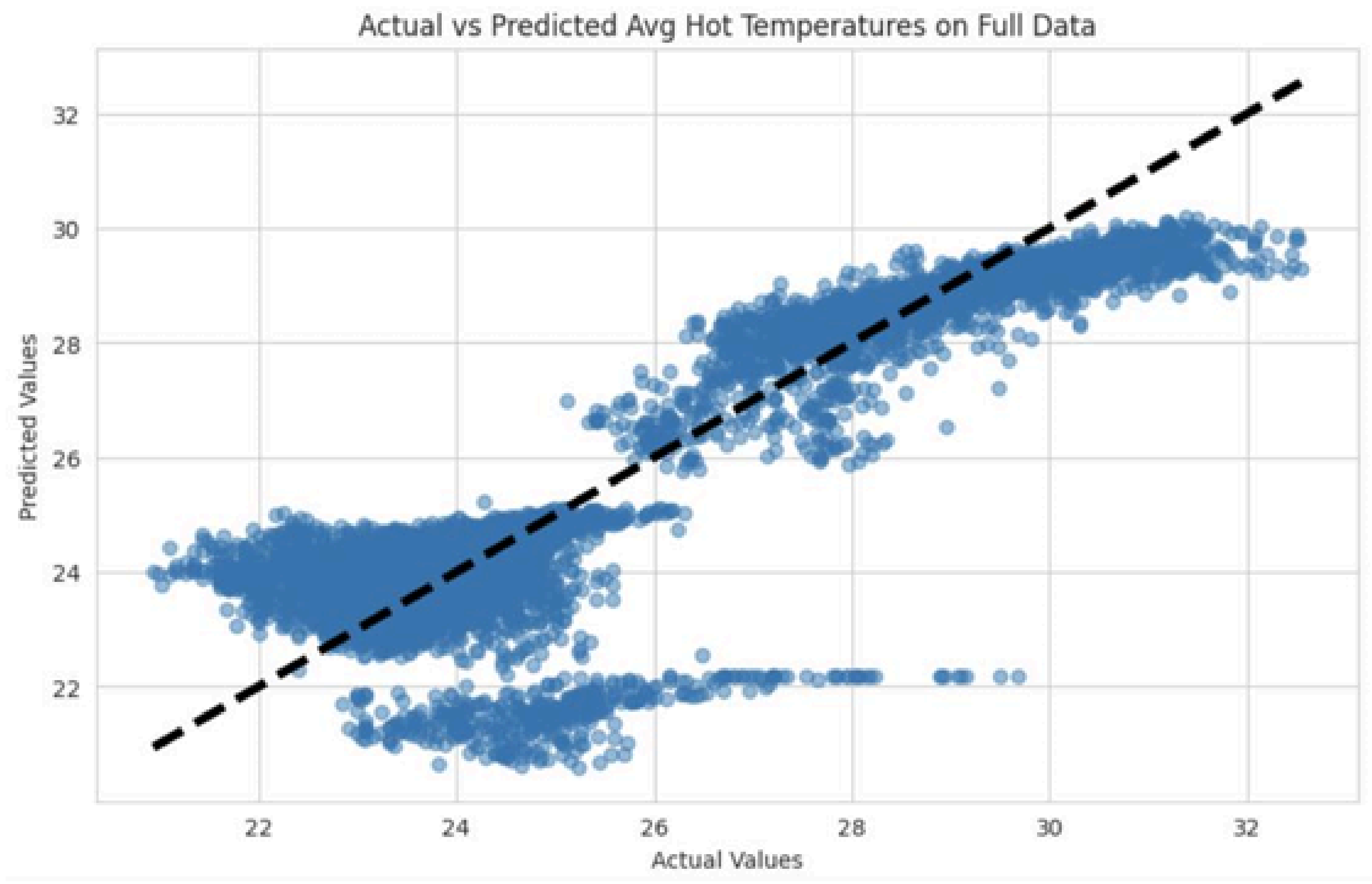
- Compares the actual and machine learning-predicted hot-aisle temperatures based on cooling parameters.
- Evaluates the effectiveness of the predictive model in optimizing thermal management.

Key Insights:

- The data points closely follow the diagonal line, indicating high prediction accuracy.
- Mean Squared Error (MSE) = 1.3156, showing the model reliably estimates temperature variations.
- Minor deviations suggest further improvements in cooling system responsiveness may be needed.

Impact:

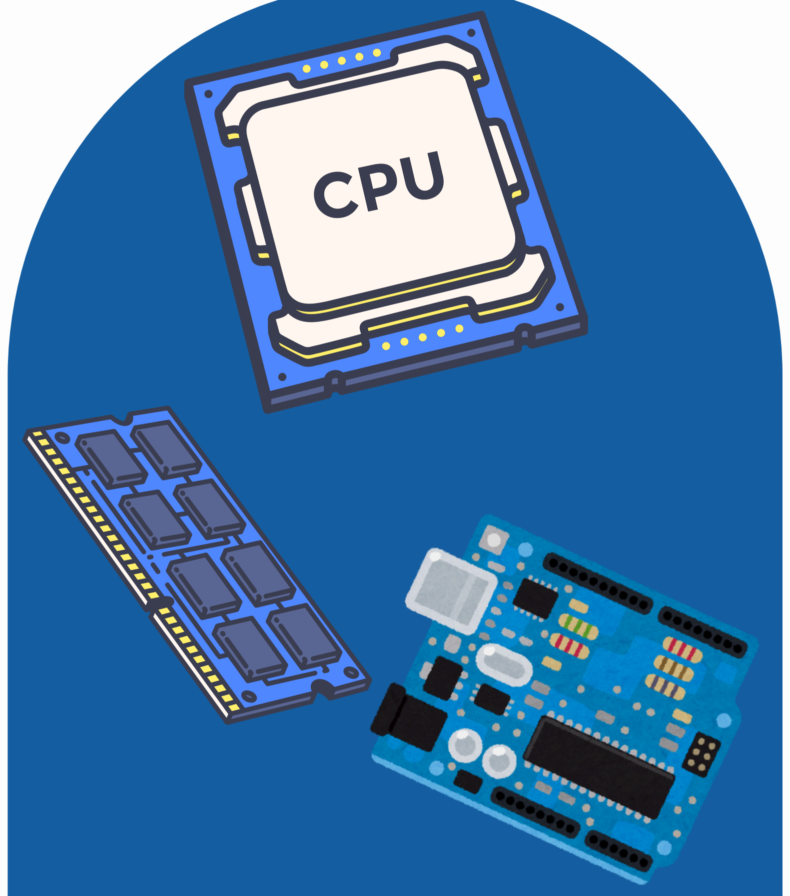
- AI-driven cooling adjustments can enhance efficiency by reducing overheating and energy waste.
- This model enables proactive thermal management, ensuring stability and sustainability in HPC data centers.



Recommendations

Strategies for Improving Energy Efficiency

- AI-driven workload scheduling: Evenly distribute computing tasks.
- Dynamic cooling automation: Machine learning to adjust fan speed and airflow.
- Real-time monitoring: Use IoT sensors to track temperature, humidity, and workload.
- Renewable energy integration: Solar/wind power for sustainable operations.

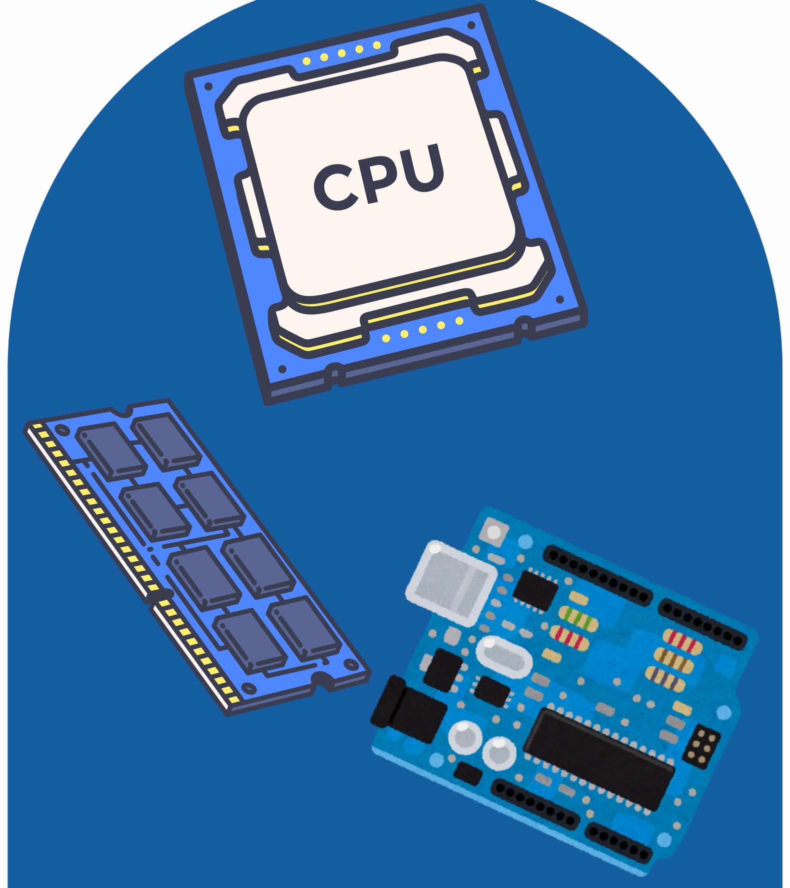


Conclusion and Future Work

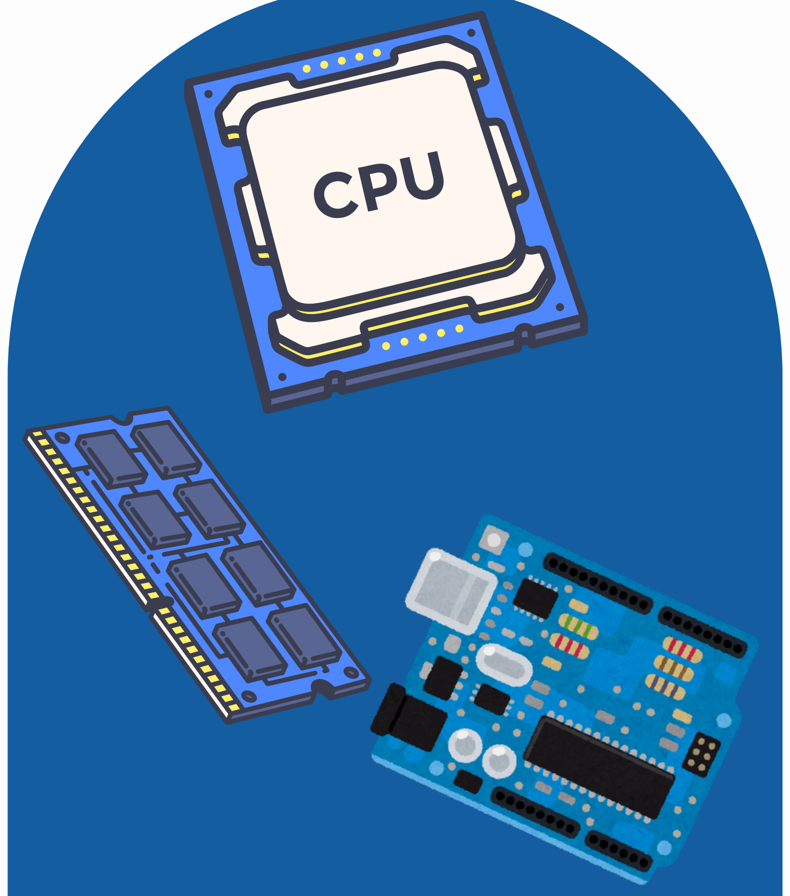
- Machine Learning optimizes data center operations efficiently.
- Proactive cooling adjustments reduce energy waste.
- Idle node shutdown enhances sustainability.

Future Work:

- Extending research to CRESCO7 and CRESCO8 clusters.
- AI-based predictive workload balancing for even greater efficiency.



Any Questions?



References

- Barroso, L.A.; Clidaras, J.; Holzle, U. The Datacenter as a Computer: An Introduction to the Design of Ware-House-Scale Machines, 2nd ed.; Morgan & Claypool Publishers: San Rafael, CA, USA, 2013.
- Mell, P.; Grance, T. The NIST Definition of Cloud Computing; National Institute of Standards and Technology: Gaithersburg, MD, USA, 2011.
- Hamilton, J. Cooperative expendable micro-slice servers (CEMS): Low cost, low power servers for internet-scale services. In Proceedings of the Conference on Innovative Data Systems Research (CIDR), Asilomar, CA, USA, 4–7 January 2009.
- ASHRAE. Thermal Guidelines for Data Processing Environments, 5th ed.; American Society of Heating, Refrigerating and Air Conditioning Engineers, Inc.: Atlanta, GA, USA, 2021; ISBN 978-1-936504-79-7.
- Hastie, T.; Tibshirani, R.; Friedman, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction; Springer Science & Business Media: New York, NY, USA, 2009.
- Georgiou, S.; Rizou, S.; Spinellis, D. Software Development Lifecycle for Energy Efficiency. ACM Comput. Surv. 2019, 52, 1–33. [CrossRef]
- Manotas, I.; Bird, C.; Zhang, R.; Shepherd, D.; Jaspan, C.; Sadowski, C.; Pollock, L.; Clause, J. An empirical study of practitioners' perspectives on green software engineering. In Proceedings of the 38th International Conference on Software Engineering, Austin, TX, USA, 14–22 May 2016.
- Bishop, C.M. Pattern Recognition and Machine Learning; Springer: Berlin/Heidelberg, Germany, 2006
- Gebreyesus, Y.; Dalton, D.; De Chiara, D.; Chinnici, M.; Chinnici, A. AI for Automating Data Center Operations: Model Explainability in the Data Centre Context Using Shapley Additive Explanations (SHAP). Electronics 2024, 13, 1628. [CrossRef]

References

- Berisha, B.; Mëziu, E.; Shabani, I. Big data analytics in Cloud computing: An overview. J. Cloud Comput. 2022, 11, 24. [CrossRef]
- Mondal, S.; Bin Faruk, F.; Rajbongshi, D.; Efaz, M.M.K.; Islam, M. GEECO: Green Data Centers for Energy Optimization and Carbon Footprint Reduction. Sustainability 2023, 15, 15249. [CrossRef]
- Hamilton, J. Cooperative expendable micro-slice servers (CEMS): Low cost, low power servers for internet-scale services. In Proceedings of the Conference on Innovative Data Systems Research (CIDR), Asilomar, CA, USA, 4–7 January 2009.
- Huerta, E.A.; Khan, A.; Davis, E.; Bushell, C.; Gropp, W.D.; Katz, D.S.; Kindratenko, V.; Koric, S.; Kramer, W.T.; McGinty, B.; et al. Convergence of artificial intelligence and high-performance computing on NSF-supported cyber-infrastructure. J. Big Data 2020, 7, 88. [CrossRef]
- Gonzalez, N.M.; Carvalho, T.C.; Miers, C.C. Cloud resource management: Towards efficient execution of large-scale scientific applications and workflows on complex infrastructures. J. Cloud Comput. 2017, 6, 13. [CrossRef]
- Panwar, S.S.; Rauthan, M.M.S.; Barthwal, V. A systematic review on effective energy utilization management strategies in cloud data centers. J. Cloud Comput. 2022, 11, 95. [CrossRef]
- Manganelli, M.; Soldati, A.; Martirano, L.; Ramakrishna, S. Strategies for Improving the Sustainability of Data Centers via Energy Mix, Energy Conservation, and Circular Energy. Sustainability 2021, 13, 6114. [CrossRef]
- Liu, C.; Yu, H. Evaluation and Optimization of a Two-Phase Liquid-Immersion Cooling System for Data Centers. Energies 2021, 14, 1395. [CrossRef]
- Çaˇglar, I.; Altılar, D.T. Look-ahead energy efficient VM allocation approach for data centers. J. Cloud Comput. 2022, 11, 11. [CrossRef]

References

- Guo,Y.; Zhao, C.; Gao, H.; Shen, C.; Fu, X. Improving Thermal Performance in Data Centers Based on Numerical Simulations. Buildings 2024, 14, 1416. [CrossRef]
- Xu,S.; Zhang, H.; Wang, Z. Thermal Management and Energy Consumption in Air, Liquid, and Free Cooling Systems for Data Centers: A Review. Energies 2023, 16, 1279. [CrossRef]
- Chen,H.; Li, D.; Wang, S.; Chen, T.; Zhong, M.; Ding, Y.; Li, Y.; Huo, X. Numerical investigation of thermal per-formance with adaptive terminal devices for cold aisle containment in data centers. Buildings 2023, 13, 268. [CrossRef]
- Wibron, E.; Ljung, A.-L.; Lundström, T.S. Computational Fluid Dynamics Modeling and Validating Experiments of Airflow in a Data Center. Energies 2018, 11, 644. [CrossRef]
- Chi, C.; Ji, K.; Song, P.; Marahatta, A.; Zhang, S.; Zhang, F.; Qiu, D.; Liu, Z. Cooperatively Improving Data Center Energy Efficiency Based on Multi-Agent Deep Reinforcement Learning. Energies 2021, 14, 2071. [CrossRef]
- Mehta, Y.; Xu, R.; Lim, B.; Wu, J.; Gao, J. A Review for Green Energy Machine Learning and AI Services. Energies 2023, 16, 5718. [CrossRef]
- Daradkeh, T.; Agarwal, A. Cloud Workload and Data Center Analytical Modeling and Optimization Using Deep Machine Learning. Network 2022, 2, 643–669. [CrossRef]
- Malik, N.; Sardaraz, M.; Tahir, M.; Shah, B.; Ali, G.; Moreira, F. Energy-Efficient Load Balancing Algorithm for Workflow Scheduling in Cloud Data Centers Using Queuing and Thresholds. Appl. Sci. 2021, 11, 5849. [CrossRef]
- Sabyasachi, A.S.; Muppala, J.K. Cost-Effective and Energy-Aware Resource Allocation in Cloud Data Centers. Electronics 2022, 11, 3639. [CrossRef]
- Grishina, A.; Chinnici, M.; Kor, A.-L.; Rondeau, E.; Georges, J.-P. A Machine Learning Solution for Data Center Thermal Characteristics Analysis. Energies 2020, 13, 4378. [CrossRef]

Thank you !