Unleashing the power of Intel CPUs in Federated learning: A comprehensive evaluation

Abstract

ASUS offers a comprehensive AI solution with a Federal AI training model and an Intel-exclusive AI algorithm to showcase how we expertly combine GPU-computing power and AI technologies.
ASUS delves deeply into the experimental design and methodology employed to assess the performance of Intel CPUs within the dynamic landscape of Federated learning (FL). Federated learning, a groundbreaking paradigm, facilitates the training of machine-learning (ML) models across decentralized devices while safeguarding local data privacy. Acting as a collaborative extension, Federated learning sets the stage for these experiments. The study places specific emphasis on evaluating the impact of AVX-512 and the oneAPI toolkit on Intel CPU performance. The implementation harnesses the Flower package for the Federated learning framework and relies on PyTorch as the backbone for ML tasks. Furthermore, the paper meticulously outlines adjustments made to Linux GRUB settings to disable AVX-512 instructions. The experimental design encompasses a thoughtfully-crafted control group with AVX-512 and oneAPI enabled, alongside several experimental groups exploring diverse configurations of AVX-512 and oneAPI settings.

Introduction

Federated learning overview

ASUS RS720Q-E11-RS8U, equipped with a 4th Gen Intel Xeon Scalable processor, is designed for efficiency, density and TCO-optimization for data centers, web servers, virtualization, cloud and hyperscale environments. RS720Q-E11-RS8U a best fit for potent data infrastructure from the cloud, to the network, to the intelligent edge. Whether you're dealing with extensive datasets or complex deep-learning models, our servers can effortlessly scale up to meet the demands of your most ambitious AI training projects. The four-node architecture design is ideal for CDN, HCI and cloud workloads to drive IT agility and digital transformation.No two AI training tasks are alike, and RS720Q-E11-RS8U acknowledges this diversity, offering the versatility to tackle a wide range of AI workloads — from image recognition to natural-language processing (NLP). Our servers are optimized to handle diverse ML models, providing a platform that adapts to the unique demands of AI-training projects. As an AI enabler, ASUS would like to deliver a comprehensive study in medical AI with Federated learning AI models.
Federated learning is an ML approach that enables training models across decentralized edge devices (such as smartphones, IoT devices or servers) without exchanging raw data. The main idea behind Federated learning is to keep the data localized, addressing privacy concerns and reducing the need to transmit sensitive information to a central server.
Here's an overview of Federated learning:
Key components:

1. Centralized server:
   - Coordinates the Federated learning process.
   - Distributes the global model to edge devices.
   - Aggregates and updates the model based on local updates from the devices.
2. Edge devices (clients):
   - Have local datasets but do not share raw data.
   - Train local models on their data.
   - Send model updates (gradients) to the central server.
3. Global model:
   - The overarching model that is trained and improved over time.

Workflow:

1. Initialization:
  · The global model is initialized on the central server.
    This model is then sent to participating edge devices.
    Local Training:
  · Edge devices train the model locally on their data.
    This training is typically done through several iterations or epochs.
2. Model update:
  · After local training, the edge device calculates the model updates (gradients) based on its local data.
3. Model aggregation:
  · The central server collects the model updates from all participating devices.
  · Aggregation methods (such as averaging) are applied to combine the updates.
4. Global model update:
  · The global model is updated using the aggregated model updates.
  · The updated global model is then sent back to the edge devices.
5. Iteration:
    Steps 2-5 are repeated for multiple iterations, gradually improving the global model.

Federated learning is a pioneering approach that facilitates the training of machine learning models across a multitude of decentralized devices. Within this paradigm, Federated learning extends the collaborative ethos, fostering cooperation within Federated learning experiments. This paper delves into the intricate evaluation of Intel CPU performance within this cutting-edge context.

Key aspects of federated learning in the medical domain

1. **Privacy preservation:**
   In the highly-sensitive realm of medical data, stringent privacy regulations necessitate innovative solutions. Federated learning emerges as a beacon, enabling collaborative model training without the need to share sensitive patient data, ensuring decentralized privacy preservation.
2. **Cross-institutional collaboration:**
   Federated learning enables seamless collaboration among medical institutions or research centers, fostering shared insights without compromising the integrity of raw patient data. This proves particularly invaluable for large-scale studies requiring diverse datasets.
3. **Real-time updates:**
   Continuous learning is paramount in the dynamic landscape of medical data. Federated learning facilitates real-time model updates, ensuring models remain pertinent and accurate as new information surfaces.
4. **Edge devices in healthcare:**
   Federated learning’s adaptability to edge devices, such as wearable sensors and implantable devices, aligns seamlessly with the increasing prevalence of Internet of Things (IoT) devices in healthcare. This approach enables personalized healthcare models without the need for continuous data transfer to a centralized server.
5. **Regulatory compliance:**
   Federated learning aligns harmoniously with regulatory frameworks like HIPAA by decentralizing data and minimizing the risk of breaches, thereby aiding healthcare organizations in adhering to stringent regulatory requirements.

Expanding on federated learning in the medical domain

The implementation of Federated learning in the medical domain introduces transformative possibilities for addressing critical challenges and capitalizing on unique opportunities:

- **Enhanced diagnostic accuracy:**
  By training models on diverse datasets from various healthcare institutions, Federated learning contributes to improved diagnostic accuracy. The collaborative nature of Federated learning ensures that models generalize well across different patient demographics and medical conditions.
- **Disease prediction and prevention:**
  Federated learning enables the development of predictive models that can anticipate disease patterns and support preventive healthcare measures. The decentralized approach allows for insights drawn from diverse datasets, leading to more robust predictions and targeted interventions.
- **Drug discovery and personalized medicine:**
  Collaborative efforts in Federated learning facilitate advancements in drug discovery by leveraging insights from multi-institutional datasets. Moreover, the patient-specific models generated through Federated learning contribute to the realization of personalized medicine, tailoring treatment strategies based on individual health profiles.
- **Efficient resource utilization:**
  In resource-limited healthcare environments, Federated learning optimizes resource utilization by allowing institutions to collectively train models without the need for centralized data storage. This approach enhances the scalability and sustainability of machine learning applications in healthcare.

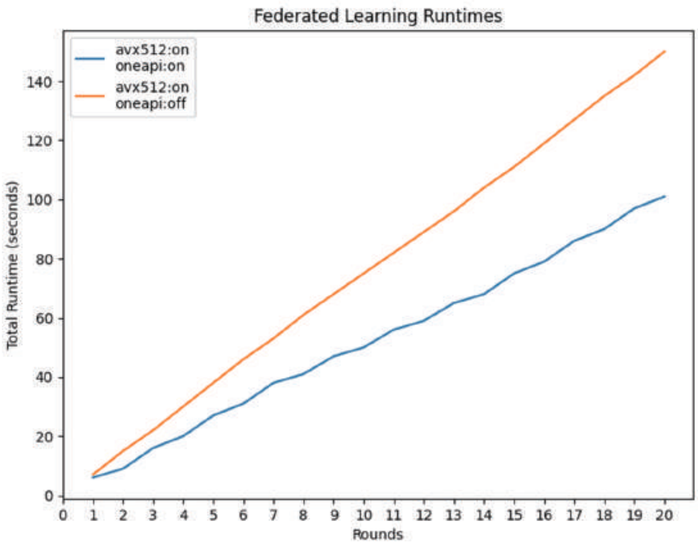
Conclusion

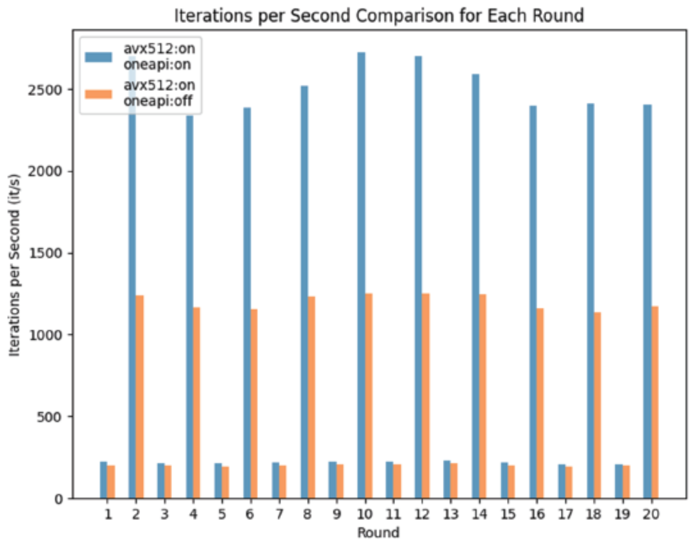
In conclusion, Federated learning, particularly in the context of the medical domain, stands out as a transformative force poised to redefine the landscape of healthcare innovation. Beyond its pivotal role in mitigating privacy concerns, Federated learning emerges as the catalyst for collaborative breakthroughs that hold the promise of reshaping critical aspects of healthcare. This groundbreaking paradigm has the potential to revolutionize diagnostics, bolster disease prevention strategies, expedite drug discovery processes, and enhance the delivery of personalized healthcare services.

The comprehensive evaluation of Intel CPU performance within this dynamic healthcare environment adds a crucial layer to the ongoing discourse surrounding the optimization of machine learning technologies. By delving into the intricacies of Intel CPUs in the context of Federated learning, this study contributes valuable insights that are poised to empower healthcare practitioners worldwide. The ultimate beneficiaries of this technological optimization are the patients, as the synergy of Federated learning and cutting-edge CPU performance seeks to usher in a new era of more efficient, precise, and accessible healthcare solutions on a global scale. This endeavor aligns with the broader mission of leveraging technology to enhance the well-being of individuals and communities, marking a significant stride forward in the pursuit of advanced, patient-centric healthcare.

Results

AVX-512:on, oneapi:off

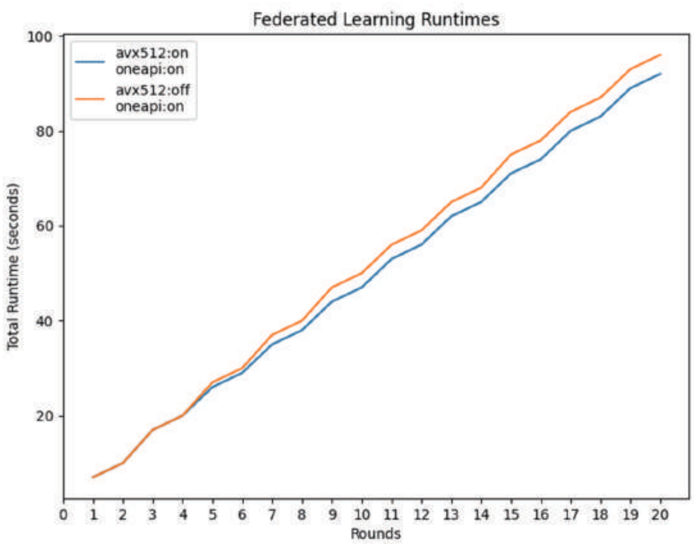


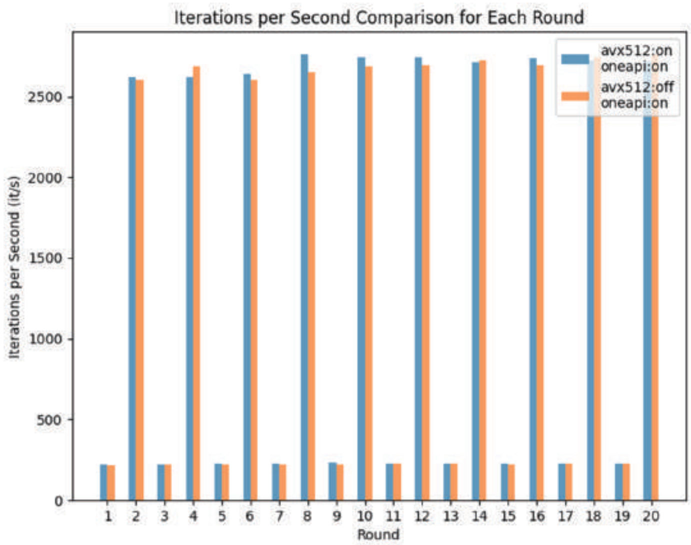


Conclusion

Enabling both AVX-512 and oneAPI in the ‘AVX-512:on, oneapi:on’ configuration results in an approximately 88.43% improvement in average iterations per second compared to the ‘AVX-512:on, oneapi:off’ configuration. This suggests that the combined use of AVX-512 and oneAPI significantly enhances the performance of Flower Federated learning on the given Intel CPU.

AVX-512:off, oneapi:on

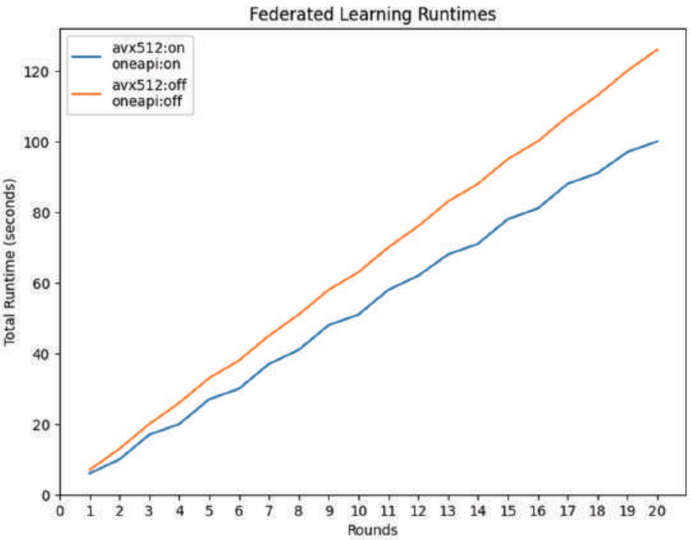


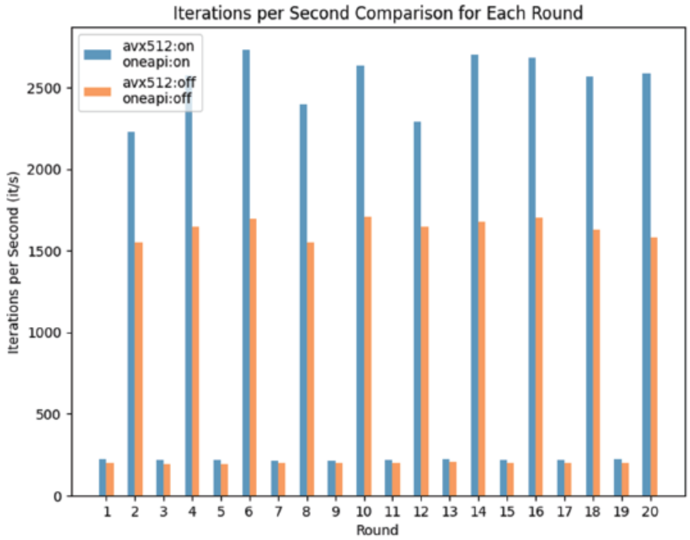


Conclusion

Configuration 2 (AVX-512 disabled, oneAPI on) shows about a 2.11% deterioration in terms of iterations per second compared to Configuration 1 (AVX-512 enabled, oneAPI on). This indicates that, in this particular scenario, enabling AVX-512 has a positive impact on performance.

AVX-512:off, oneapi:off





Conclusion

The ‘AVX-512:on, oneapi:on’ configuration demonstrates a roughly 33.77% improvement in iteration speed over the “AVX-512:off, oneapi:off” configuration across the 20 rounds of Flower Federated learning. Therefore, enabling AVX-512 and oneAPI appears to significantly enhance the performance of Flower Federated learning on Intel CPUs.

Why AVX-512, oneAPI are important to AI training ?

In summary, when deploying Flower Federated learning on an Intel CPU, the optimal configuration emerges as one that harnesses the combined power of both AVX-512 and oneAPI. Notably, the performance consistently outpaces configurations where either AVX-512 or oneAPI is disabled. This highlights a synergistic relationship between these two advanced features, demonstrating that their simultaneous activation unleashes the full potential of the Intel CPU in the context of Flower Federated learning.

Enabling AVX-512, which leverages advanced vector processing capabilities, and integrating oneAPI, a comprehensive toolkit designed for parallel programming, creates a harmonious environment for efficient and accelerated machine learning tasks. The observed superior performance underscores the importance of a holistic approach, where these features complement each other, resulting in optimized computational power and streamlined processing for Flower Federated learning.

This finding not only substantiates the significance of a unified configuration but also underscores the adaptability and efficiency of Flower Federated learning when running on Intel CPUs. By unlocking the combined capabilities of AVX-512 and oneAPI, users can harness a level of performance that transcends the sum of its parts, marking a pivotal advancement in the realm of Federated learning on Intel architecture. This optimized configuration promises to be a game-changer for practitioners seeking to maximize the computational prowess of their Intel CPU, ultimately enhancing the overall efficiency and efficacy of Flower Federated learning in diverse ML scenarios.

Experimental setup

Server configuration

- **Operating system:** Debian 12
- **CPU:** 8-core Intel Xeon Silver 4410T
- **RAM:** 8192 MB
- **Software:**
    - Miniconda Python 3.9
    - Flower package (Flower GitHub)

Intel oneAPI toolkit installation

Detailed instructions for installing the oneAPI toolkit are available in the Intel oneAPI documentation.

Client configuration

- **Operating system:** Debian 12
- **CPU**: 8-core Intel Xeon Silver 4410T
- **RAM:** 8192 MB
- **Software:**
    - Miniconda Python 3.9
    - Flower package
    - Intel oneAPI Toolkit (installed using provided shell script)
- **AVX-512 control:** Via Linux GRUB settings

Flower framework setup

Utilizing the Flower package for FL experiments, the PyTorch-based examples (Flower GitHub - PyTorch Examples) are configured to run with 10 rounds of testing.

AVX-512 and oneAPI configuration

Control group

- **Client1:** AVX-512 enabled, oneAPI enabled

Experimental groups

- **Client2:** AVX-512 enabled, oneAPI disabled
- **Client3:** AVX-512 disabled, oneAPI enabled
- **Client4:** AVX-512 disabled, oneAPI disabled

Linux GRUB settings modification

To disable AVX-512 instructions, modify the /etc/default/grub file according to Intel oneAPI documentation.