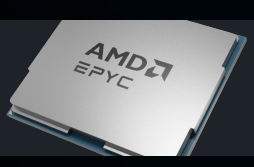
4TH GEN AMD EPYC™ 9004 PROCESSORS
ADVANCE ENTERPRISE AI INFERENCE

AMD LEADERSHIP ENTERPRISE AI PORTFOLIO HELPS CONSOLIDATE TRADITIONAL & AI WORKLOADS

Key features enable AMD EPYC to consolidate infrastructure, optimize costs, and adapt to evolving needs of both traditional and AI workloads.

- **Advanced Memory Management**
- **Robust Software Ecosystem**
- **Hardware Acceleration**
- **High Core Count & Multithreading**
- **Large Cache Sizes**

AMD
Offers Leadership
Performance & Efficiency
for both CPUs & GPUs



AMD EPYC

- Mixed workload Inference
- Small to Medium Models
- Batch/Small Scale Inference

AMD INSTINCT

- AI Training & Dedicated Deployments
- Medium to Large Models
- Large-Scale Inference

LEADERSHIP ENTERPRISE AI CPU INFERENCE & ENERGY EFFICIENCY SOLUTIONS

AMD EPYC processors deliver incredible performance for AI inference, and the optimized AMD ZenDNN software library helps deliver even greater performance gains.

~68% Higher Image Classification Throughput

YOLO v5 images/sec BF16 precision more is better, batch size=960

ZD-052

~96.8% Faster Image Recognition

ResNet50 latency at BF16 precision lower is better, batch size=1

ZD-053

~36% Faster Natural Language Processing

Llama2 13B latency at FP32 precision lower is better, batch size=1

ZD-053

2P AMD EPYC 9654 powered server comparing throughput performance of select AI benchmarks running the ZenDNN Plugin for PyTorch 4.2.0 (zentorch) to performance on Native PyTorch

AMD EPYC Processors Power the World's Most Energy-Efficient Servers EPYC-028D

up to **1.7x**

Higher Average Inference FPS per CPU Watt Running Select Published Phoronix Test Suite OpenVINO™ Workloads

SP5-252

2.25x

Overall ssj_ops per Watt Running SPECpower_ssj®2008

SP5-011F

2P 4th Gen AMD EPYC 9754 (128C) vs. 2P 5th Gen Xeon 8592+ (64C) Powered Servers

DEPLOY WITH THE CONFIDENCE OF ADVANCED SECURITY FEATURES & OPEN STANDARDS

Compute with confidence, knowing that your business is addressing today's security challenges with the advanced security features of AMD Infinity Guard.[1]

Plus, long and consistent AMD commitments to supporting open standards is critical to the development of a healthy and competitive computing ecosystem.

"We have found that AMD has the most powerful processor on the market and that helps us build systems that can increase the throughput for each server while reducing the hardware investment costs."

Professor Minh Hoai Nguyen
Principal Research Scientist & Head of Smart Edge, VinAI

Case Study: https://www.amd.com/en/resources/case-studies/vin-ai.html

1. AMD Infinity Guard features vary by EPYC™ Processor generations and/or series. Infinity Guard security features must be enabled by server OEMs and/or Cloud Service Providers to operate. Check with your OEM or provider to confirm support of these features. Learn more about Infinity Guard at https://www.amd.com/en/technologies/infinity-guard. GD-183A

For details on the claims used in this document, visit amd.com/en/legal/claims/epyc.

PID: 242806105-A August 2024