Expert Community

Intel Demo session
11 09 2019 – 18.00

Demonstrate running MXNet and Deep recommendation engine utilizing Intel DL Boost

Pawel Gepner
Field Application Engineer – Intel Corporation

Trusted partner for your **Digital Journey**

Atos

Demonstrate running MXNet and Deep recommendation engine utilizing Intel DL Boost

Majority of commercial deep learning applications today use 32-bits of floating-point precision (FP32) for training and inference workloads. However different scientists have demonstrated that both training and inference can be performed with lower numerical precision, consequently, the use of lower-precision operations is expected to become standard practice soon. In April 2019, Intel announced the 2nd generation of Intel Xeon Scalable processors (codenamed Cascade Lake) with Intel Deep Learning Boost technology. Intel Deep Learning (DL) Boost technology includes integer vector neural network instructions (VNNI), branded as Intel DL Boost, accelerating numerous of AI training and inference workloads, increasing the performance by 2–3x compared with first generation of Intel Xeon Scalable processor. In this talk we look what the 2nd generation of Intel Xeon Scalable processors with Intel Deep Learning Boost technology, brings to artificially intelligence (AI) type of workloads. In addition we will discuss the prospect of new ISA extension BF16 and potential benefits of these extensions.

Pawel Gepner

Field Application Engineer – Intel Corporation

Paweł Gepner
Field Application Engineer focused on High Performance Computing platforms at Intel France.

Pawel joined Intel in 1996 as Field Application Engineer for Central and Eastern Europe.
In 2001, he became EMEA Architect focused on HPC area. Currently he is responsible for the development and design of future server platforms and HPC systems at one of Intel's leading EMEA partners.

Pawel is the Intel Corporation spoke person responsible for communication with the press reg. technical and technology aspect of Intel's products and technology.
He led server development projects including first Fault Tolerance Systems based on IA-32 from Stratus Technology. and was responsible for driving Pentium III server project at IBM Development Center in Greenock. Pawel also led the team of Intel architects that developed Bull's Itanium 2 system. He was also involved in Itanium 2 projects at Siemens AG and Eriksson. He led the development team for the first teraflop computing projects in EMEA and first Itanium 2 teraflop installations. He was driving many of the HPC projects in including TASK, SKODA, VW, CERN, and many others.

Pawel graduated in Computer Science and he holds master's and Ph.D. degrees from Warsaw University of Technology, Poland and habilitation from Czestochowa University of Technology - Associated Professor degree.
He has written 50 technical papers on Computer Science and Technology and is also a board member and technology advisor for many international scientific and commercial HPC projects.