XILINX

Accelerated Machine Learning Suite for AWS f1 and Alveo

- **Up to 12x speedup for machine learning algorithms**
- **Seamless integration for Apache Spark™**
- **Lower OpEx by 2x**
- **Available on Cloud (AWS f1) and on-premise (Alveo)**

INTRODUCTION

Apache Spark™ is one of the most widely used frameworks for data analytics and machine learning. Spark has been adopted widely in recent years for big data analysis by providing a fault-tolerant, scalable and easy to use in-memory abstraction. Spark Mllib is widely used for machine learning applications on large data sets.

InAccel provides ready-to-use machine learning IP cores for widely-used algorithms for classification, clustering and recommendation applications. InAccel's IP cores can be used to speedup the applications and reduce the OpEx without code modifications.

PRODUCT OVERVIEW

FPGA-based Machine learning Accelerators available on:

- Cloud (Amazon AWS f1.x2 .x4 .x16, Nimbix)
- On-premises (Xilinx Alveo Data Center Accelerator Card)

SOLUTION OVERVIEW

FPGA-based ML suite for Apache Spark

InAccel provides FPGA-Accelerated ML suite for distributed systems as a fully integrated AMI/AFI that in used to speedup machine learning applications for Apache Spark™. The suite provides all the required APIs that allows the utilization of the Amazon F1 instances without any changes of the original code.

InAccel provides both the IP cores and a novel scalable resource manager that allows the scalability of the FPGA instances in the cloud or on-premise data centers utilizing Xilinx FPGA cards. InAccel accelerated ML-suite offers up to 3x speedup compared to multi-threaded contemporary processors and 2x lower OpEx.

3x-12x Speedup

2x Lower Cost

Zero-code changes

Available IP cores:

Classification: Logistic regression (BGD)
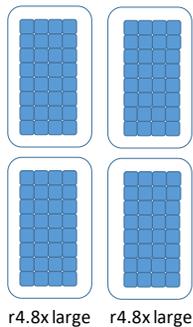
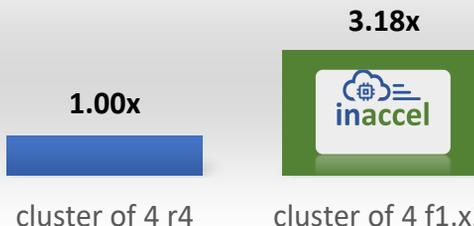Clustering: K-means

Recommendation engines (ALS-based)

Cluster of 4 r4.8xlarge
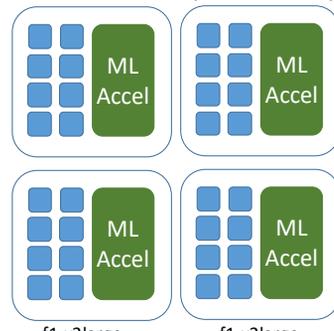

r4.8x large   r4.8x large

| 128 cores |

Speedup comparison for a cluster of 4 nodes

3.18x

1.00x

cluster of 4 r4   cluster of 4 f1.x2

| $2.1 x 1 = $2.1/hour | $3.0 x 1/3.2 = $0.9/hour |

Cluster of 4 f1x2 (SW + InAccel)

ML Accel   ML Accel   ML Accel   ML Accel

f1.x2large   f1.x2large

| 32 cores + InAccel IPs |

Benchmark results

Comparison between a Spark cluster of 4 nodes. The SW only solution is based on a cluster of r4.8xlarge instances ($2.1/hour) and the InAccel cluster uses 4 f1.x2large instances ($1.6/hour for f1.x2 and $1.4/hour for InAccel cores). The benchmark is based on logistic regression over Spark 2.0. The InAccel Accelerated ML suite allows 3.18x speedup and allows up to 2x reductions on the OpEx compared to the r4.x8 instances. Compared to f1 SW-only, it achieves 12x speedup.

OS:	Linux/Unix, Ubuntu 16.04
Price:	
f1.x2	$1.4/hour or $500/month ($0.7/hour)
f1.x16	$10/hour or $3000/month ($4.2/hour)

CONCLUSION

Use now our Accelerated ML suite to speedup your applications without any changes on your code. Use InAccel IP cores to speedup your clustering, classification and collaborative filtering applications. Contact us if you need more specialized solutions.

Supported APIs

TAKE THE NEXT STEP

Try now for free the InAccel Accelerated Machine Learning Suite for Spark, C/C++, Python , Java or Scala on www.inaccel.com or in AWS marketplace.

Corporate Headquarters

InAccel, Inc.
500 Delaware Ave, STE 1, #1960
Wilmington,
Delaware, 19 801
U.S.A.

Europe

InAccel Hellas
Formiono s 47
Kaisariani, 16121,
Athens
Greece

www.inaccel.com

info@inaccel.com

/inaccel

@inaccel