SOLUTION BRIEF

TigerGraph: Cosine Similarity Acceleration

INTRODUCTION

Healthcare, financial services and manufacturing businesses demand faster, deeper, and wider insights from their connected data, and TigerGraph's proven graph platform technology connects data silos for deeper, wider and operational analytics at scale.

The massive parallel processing capability of Xilinx Alveo accelerator cards delivers TigerGraph cosine similarity search results ***300 times faster*** than a CPU-based approach.

SOLUTION OVERVIEW

Better and faster data insights create better outcomes: from improving critical patient care in healthcare to understanding customer needs and recommending the right product in retail. However, in traditional relational or NoSQL databases, this data can be hard to extract and slow to obtain. TigerGraph leading scale-out graph database provides a complete, distributed, parallel graph computing platform supporting web-scale data analytics in real-time.

TigerGraph, accelerated with Xilinx Alveo U50 cards and based on the massively parallel processing capability of FPGA architecture, offers superior results for computing cosine similarity calculations.

Features and Benefits

Full FPGA accelerated pipeline

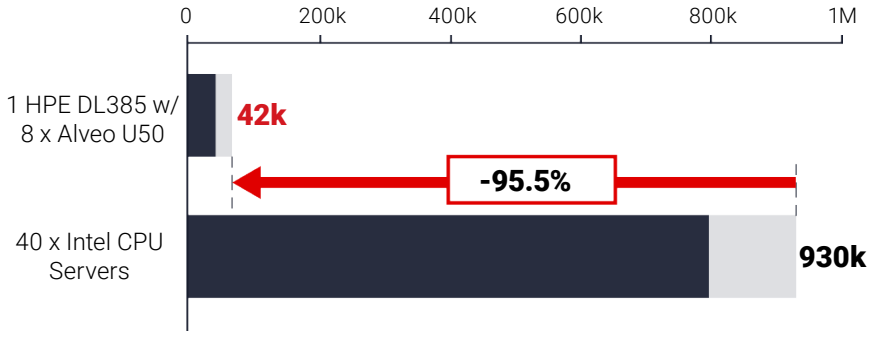
- Faster, deeper, and wider insights on connected data
- 300x better performance vs CPU
- 95% lower TCO vs CPU
- Native parallel graph database supports loading terabytes of data
- TigerGraph GSQL transparent acceleration
- Ready-to-use Xilinx Cosine Starter Kit for use in Healthcare and Retail applications

Conventional predictive modeling for clinical outcome prediction fails to deliver high accuracy for patients with unique characteristics. Using TigerGraph's cosine similarity algorithm, providers can dynamically personalize clinical outcome prediction by identifying and analyzing similar past patients rather than all available patient records.

This improved prediction performance comes with an increased computational burden, along with attendant increased costs for CPU-based deployments.

For cosine similarity acceleration on an 80 million patient database, Xilinx Alveo accelerator cards deliver huge improvements in both CAPEX and OPEX.

Eight Xilinx Alveo U50s, mounted in a single 2U HPE DL385 Gen10 Plus server featuring industry-leading EPYC CPUs, delivers the same performance as 40 Intel CPU based servers, with 95% lower TCO.



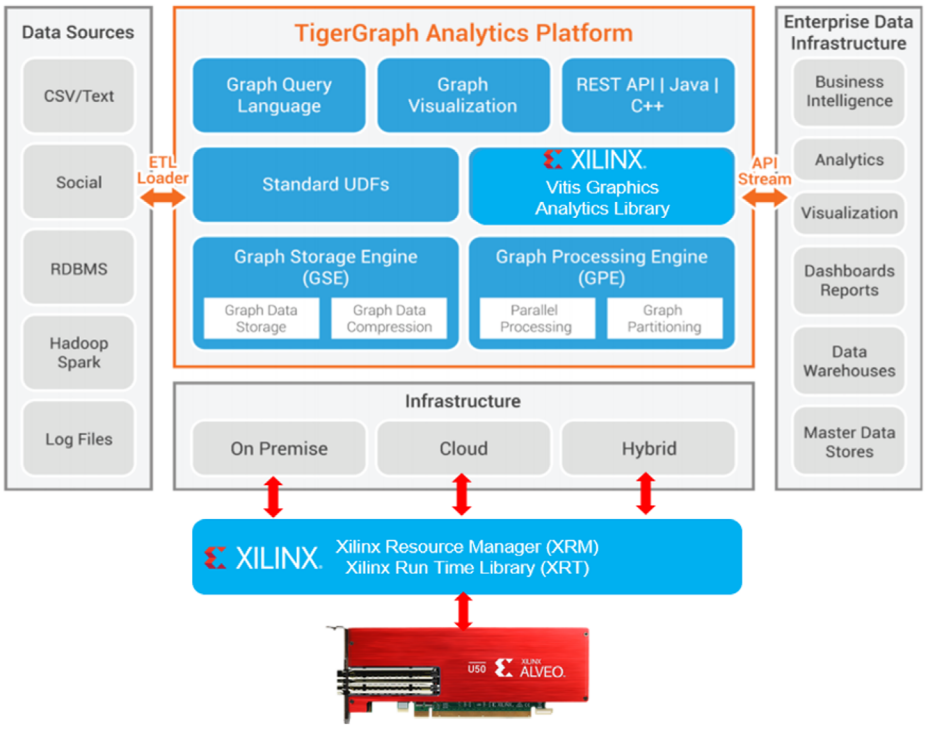
Adaptable. Intelligent.

SOLUTION DETAILS

The **Xilinx Vitis Graph Analytics Acceleration Library** is seamlessly integrated with TigerGraph as a plugin. The plugin consists of two major components: the Xilinx Cosine Sim Library and the Xilinx Runtime and Resource Manager.

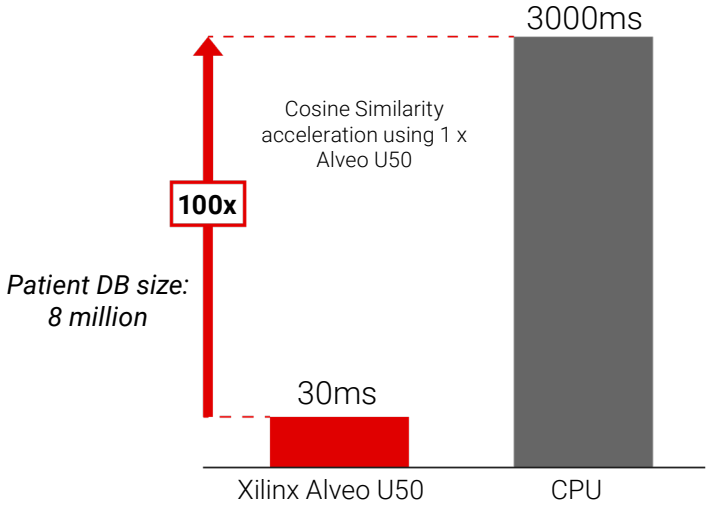
The **Xilinx Cosine Sim Library** is provided as a custom TigerGraph User Defined Function (UDF) that can be called directly from GSQL.

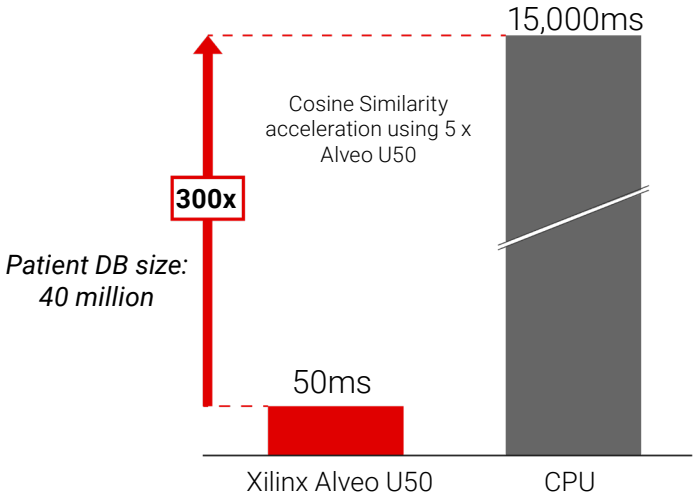
The **Xilinx Runtime and Resource Management** libraries are installed on all processing nodes to manage resource allocation on Alveo Acceleration cards and data movement between the CPU and the FPGA.



PERFORMANCE

As the size of database increases, the FPGA-based Alveo acceleration scales linearly in performance versus a CPU-based implementation. For the CPU implementation, memory bandwidth between x86 system memory (DDR) and CPU cores is limited. When multiple CPU cores work simultaneously, the system stalls as data cannot move fast enough from DDR. When this limitation is reached, scaling more CPU cores only increases TCO costs without any corresponding increase in performance.





Notes:
Dataset: Synthetic patient data generated by "Synthea" (https://synthetichealth.github.io/synthea/)
Algorithm: Cosine Similarity (cos theta between property vectors)
Property Vector: 197 int properties for each customer (patient)
CPU: 2x Intel(R) Xeon(R) Silver 4116 CPU @ 2.10GHz, 48 cores [100% utilized], 528 GB RAM card

TAKE THE NEXT STEP > Visit www.xilinx.com/tigergraph

XILINX

Adaptable. Intelligent.