

Xelera Silva

INTRODUCTION

Gradient Boosting frameworks such as XGBoost, LightGBM and CatBoost, as well as Random Forest algorithms are widely used in ransomware and DDOS detection systems, recommender systems and trading systems. **Xelera Silva** – a Decision Tree acceleration software – provides best-in-class throughput and latency for XGBoost, LightGBM and Random Forest inference by leveraging COTS datacenter-grade FPGA accelerators.

NETWORK THREAT DETECTION

The network is the first instance that must be protected in order to secure the enterprise IT system. Machine learning-powered firewalls perform application-level inspection to detect suspicious patterns in the network flow data (e.g. in the case of a DDOS attack). **Silva** provides the inferencing speed required to ensure that computationally demanding machine learning-based firewalls can operate at an adequate speed to avoid a slowing-down of the IT system.

RANSOMWARE DETECTION

Ransomware cyber attacks are rising at a rapid pace and traditional security approaches often fail to intercept them. Gradient Boosting Decision Tree and Random Forest algorithms are among the most widely used techniques for the automated detection of ransomware attacks. A high filtering throughput and low detection latency are the key objectives in these systems. **Silva** satisfies these performance constraints by ensuring a high throughput at an ultra-low latency.

SOLUTION BRIEF

XELERA

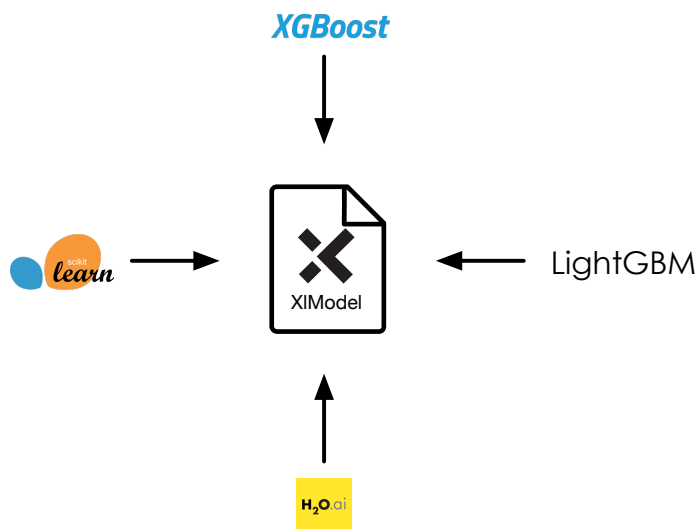
ANALYTICS

- Fastest XGBoost, LightGBM and Random Forest inference
- 10x higher throughput compared to CPUs and GPUs
- Ultra-low, guaranteed and controllable latency
- Usable with zero code change

OVERVIEW

Xelera Silva speeds up the inference of Gradient Boosting Decision Tree and Random Forest algorithms. It works with models created with **XGBoost**, **LightGBM**, **Scikit Learn**, and **H2O.ai**.

- **Acceleration:** 10x higher throughput than GPUs, 50x higher throughput than CPUs
- **Guaranteed latency:** Milliseconds query latency with 0.1ms jitter
- **Seamless adoption:** Usable with standard machine learning frameworks in a bring-your-own-model fashion, usable with zero code change
- **Deployment:** Built to run out-of-the-box on COTS server and datacenter FPGA hardware, and on public cloud instances



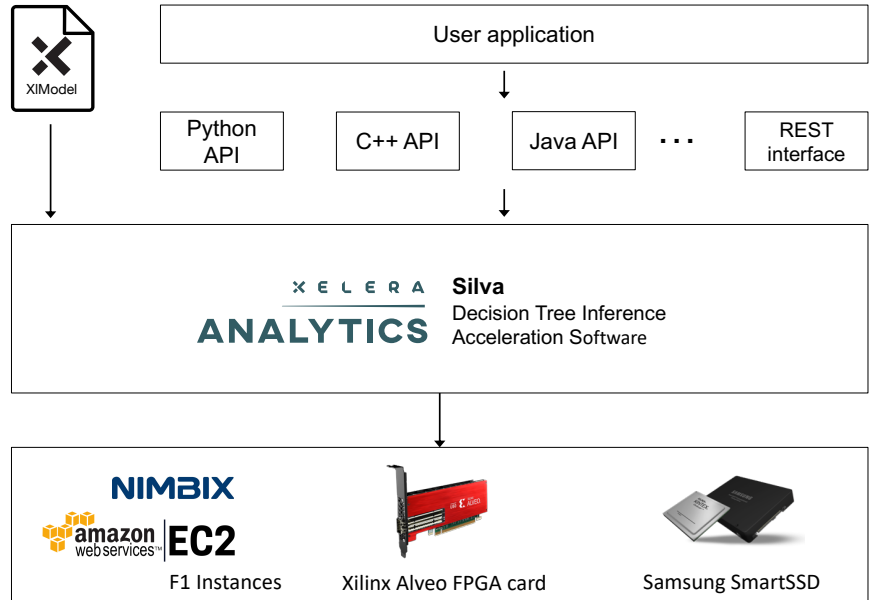
H2O® is a trademark of H2O.ai.

Xelera Silva

USAGE DETAILS

The **Xelera Silva** acceleration software consists of two parts:

1. The **Model Compiler** converts the machine learning models from **XGBoost, LightGBM, Scikit Learn,** and **H2O.ai** automatically into a unified model format (XIModel).
2. The Decision Tree acceleration software executes the compiled XIModel on Xilinx Alveo platforms, AWS F1 and Nimblex cloud instances, or the FPGA-enhanced Samsung SmartSSD. The user application interacts with the acceleration software via a Python, Java, C++ or REST interface. The interface is compatible with the above machine learning frameworks. The platform is discovered automatically.

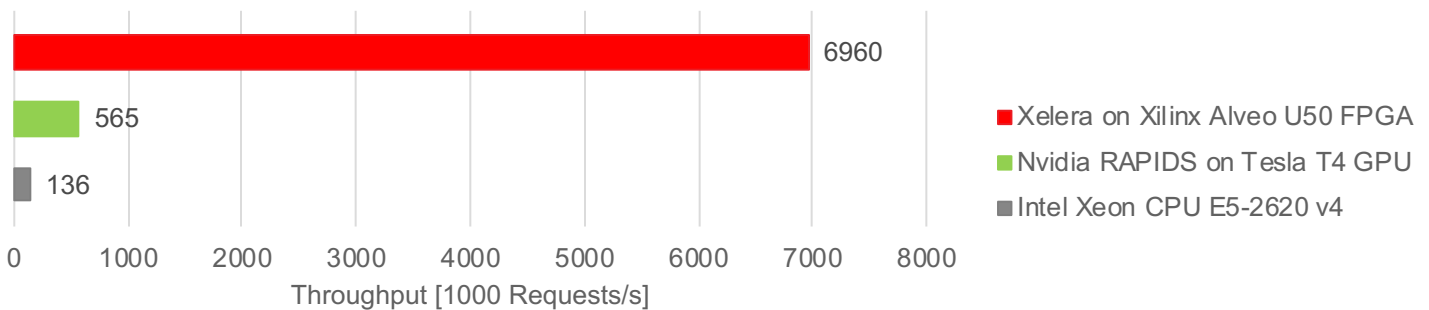


BENCHMARK

The graph below shows a throughput comparison for an **XGBoost** regression between FPGA, GPU and CPU platforms, measured using a publicly available data set. The benchmark is exemplary in that the throughput depends on the parameters of the input data set, the algorithm, and the machine learning model.

Data set: Malware Detection (Kaggle) (<https://www.kaggle.com/c/malware-detection/overview>) | Number of features: 23 | Maximum tree depth: 10 levels | Number of trees: 200

- FPGA acceleration software: Xelera Silva for XGBoost
- GPU acceleration software: Nvidia RAPIDS CuML
- CPU: XGBoost (no hardware acceleration)



TAKE THE NEXT STEP

Request a free trial (on-premises or public cloud): <https://xelera.io/product/demo-license-requests>

05 January 2021