

Alveo Ecosystem Partner Solutions

Oct 10, 2018



Meet ALVEO™





U200

892K
LUTs

35MB
Internal SRAM
Capacity

31TB/s
Internal SRAM
Bandwidth

3100img/s
CNN Throughput*

U250

1,341K
LUTs

54MB
Internal SRAM
Capacity

38TB/s
Internal SRAM
Bandwidth

4100img/s
CNN Throughput*

*Low-latency GoogLeNet v1

Alveo Ecosystem



In-Memory NoSQL Database

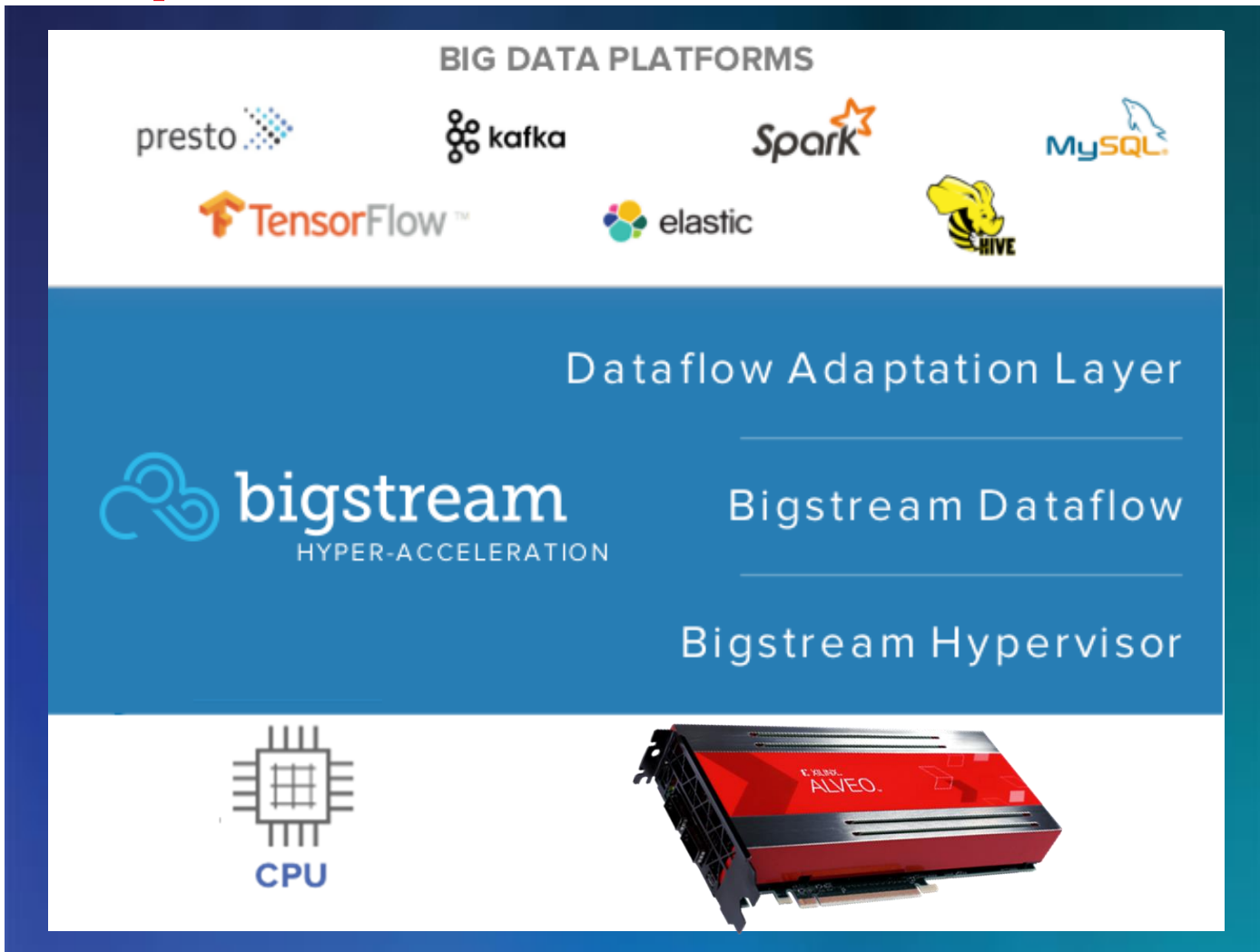


Examples:	Key	Value
Directory	Company Algo-Logic	Phone # (408) 707-3740
Forwarding Tables	IP Address 204.2.34.5	Interface : MAC Address Eth6 : 02:33:29:F2:AB:CC
Data De-duplication	Content Hash XYZ	Storage Block ID 948830038411
Stock Trading	Order ID ATY1121791101	Symbol, Side, Price AAPL, B, 126.75
Graph Search	Virtex v140	Edge List v201, v206, v225

- > Up to 170 Million searches per second
- > Under 400 ns Latency
- > Table depth of up to 12 Million records per Alveo U200 Data Center Acceleration Card
- > Client software API compatible with C/C++, Java, Python, and other programming languages
- > 100-1000x speedup vs CPU

Stock Market Order IDs | Pattern matching | Sensor Capture | Control Systems

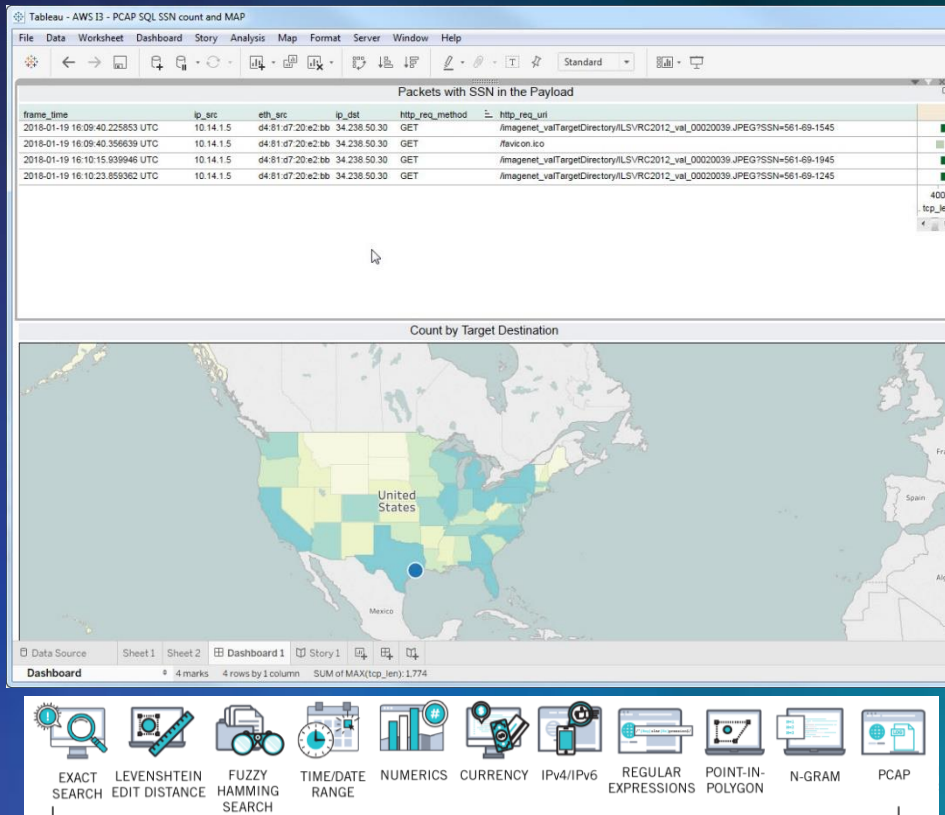
Spark Acceleration



- > Zero code change
- > Cross platform – intelligent, automatic computation slicing across CPUs and Alveo Datacenter accelerators
- > 2x- 30x acceleration

Spark Acceleration: ETL | Streaming | Kafka | Hyve | Spark SQL

High Speed Search Acceleration – Integrated with Tableau

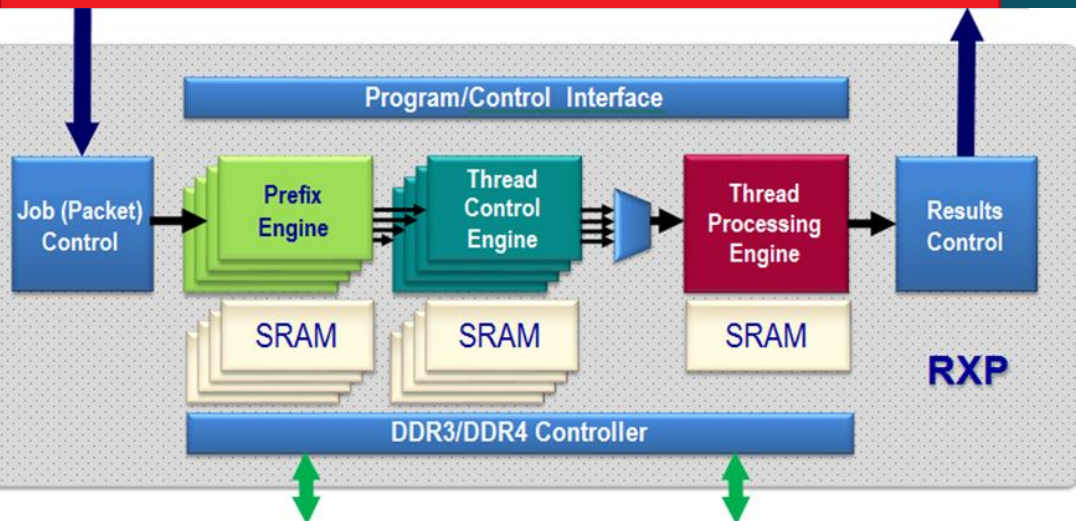


- > Near real-time search performance
- > Add complex queries including fuzzy search, PCAP analysis, and regular expression capabilities
- > Eliminate ETL/indexing for fast, varied data (XML, JSON, CSV, Unstructured, PCAP)
- > Up to 90x acceleration vs CPU

Accelerate: Splunk | Elastic Search | Microsoft Excel

High-Speed Regular Expression Processor (RXP)

Xilinx SDAccel™ Framework



- > Enhanced security analytics acceleration
- > Flexible platform for high throughput/high rule density network\
- > Easy to build and configure modular security or data analytics products on FPGA
- > Titan's RXP generates a 10X+ improvement in throughput and can support up to 1 million rules to achieve significant CPU offload and free up network resources

RegX Optimum acceleration up to 100Gb/s

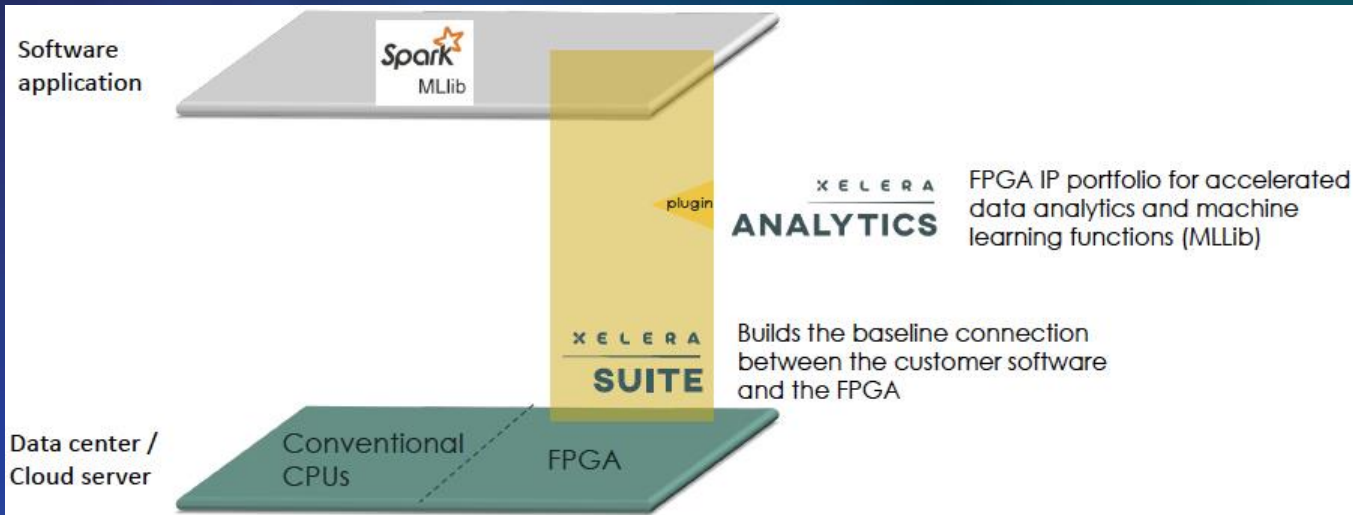


- > 100% compatibility with the open source GPDB project.
- > On average 3x faster queries than Greenplum DB when running TPC-H queries.
- > Snippets of GO or PYTHON code can be interweaved into SQL queries to group and send the records to TensorFlow or other ML libraries for training or inferencing

Deepgreen DB Acceleration | Geospatial Join | In-Database Machine Learning

Acceleration for Apache Spark MLlib

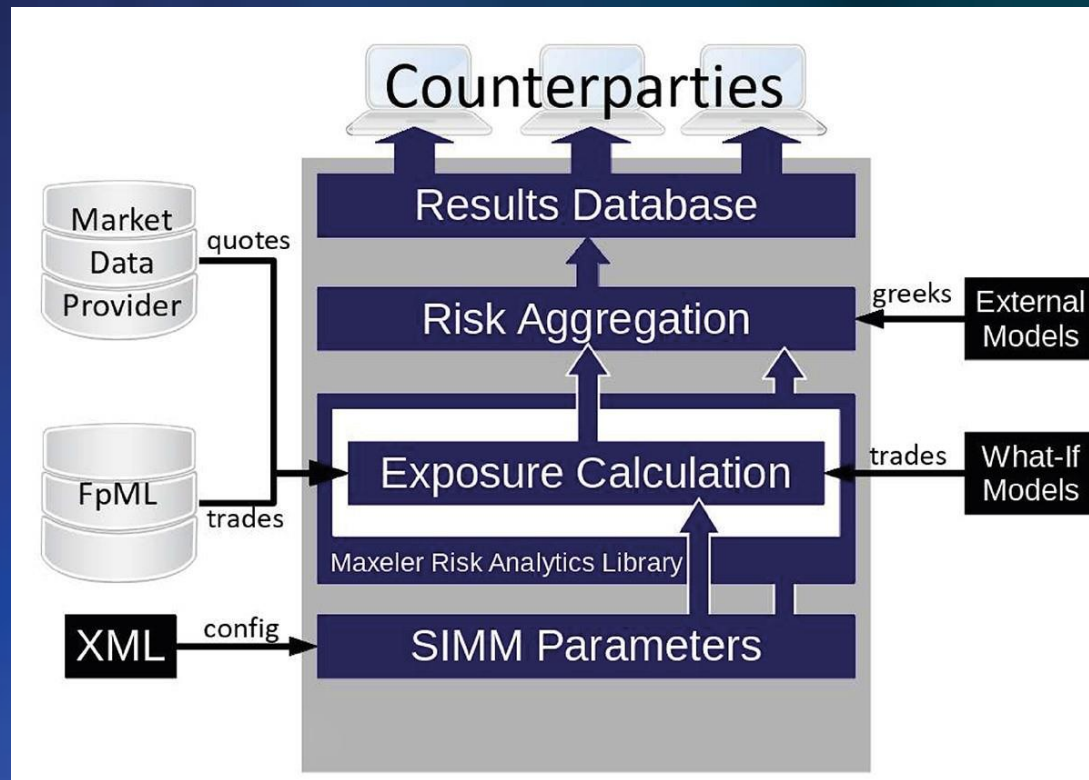
X E L E R A



- > **Xelera Analytics enables acceleration for Apache Spark MLlib**
- > **Fully integrated into Spark framework**
- > **25x speed-up vs. CPU**

Plug-and-play acceleration | Advanced data analytics | real-time business applications

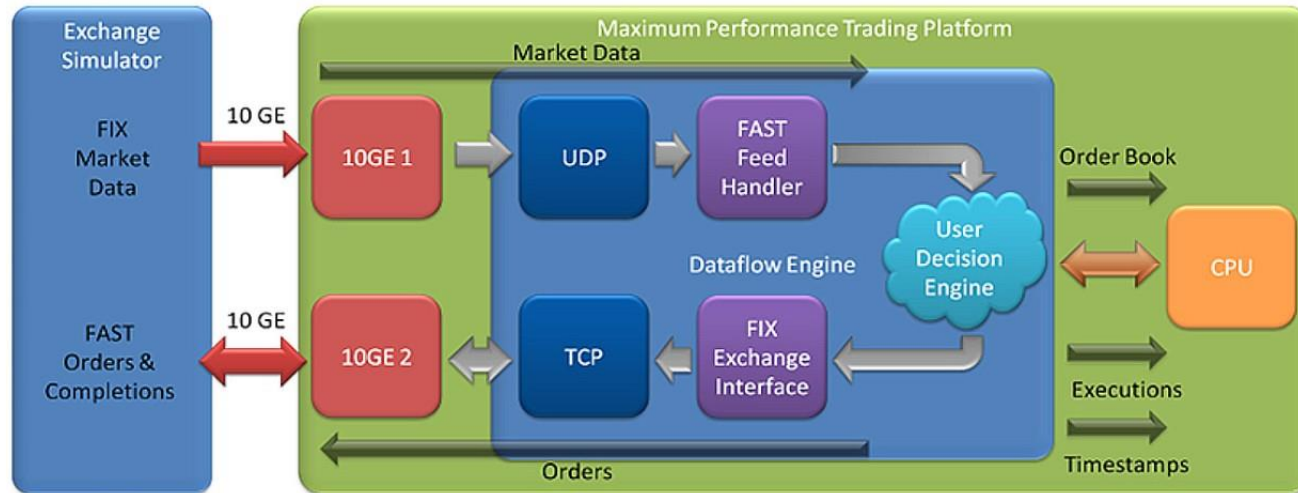
Standard Initial Margin Model (SIMM)



- > Industry-proven risk analytics infrastructure.
- > Performance advantage for complex risk calculations of 10-50x for real-time results before trading
- > Initial margin requirements calculated directly from a portfolio of trades

Automatically select best counterparty | Evaluating “what-if” trades and hedges

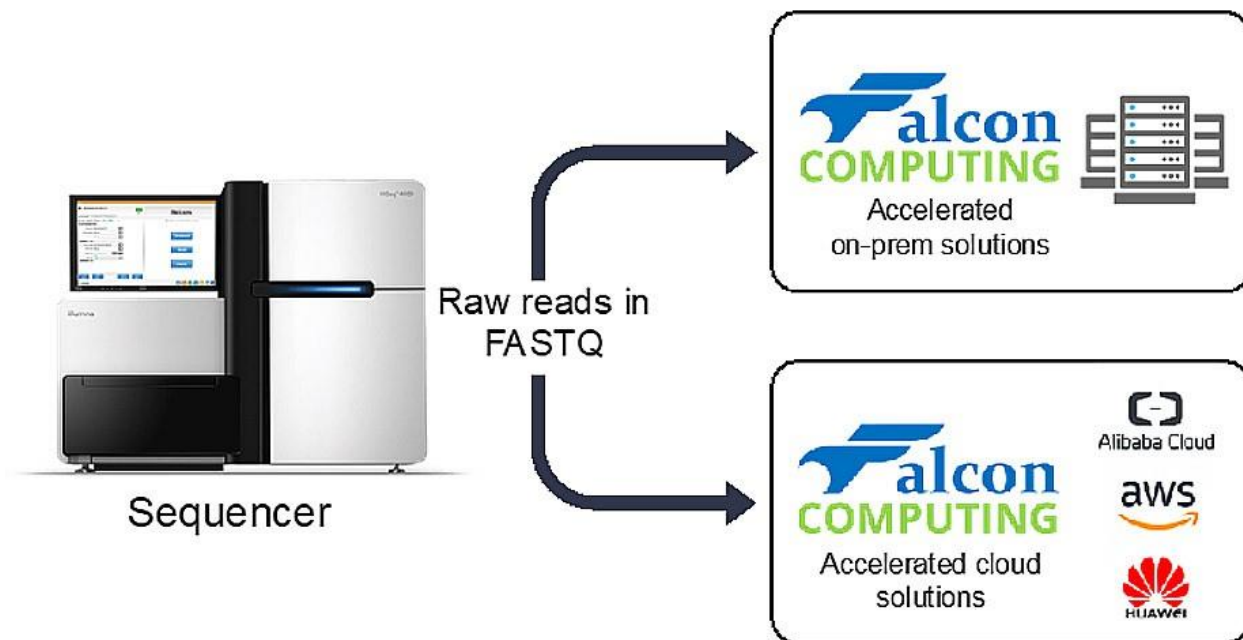
Real Time Risk Dashboard



- > Real-time risk enables pre-trade computations
- > Dashboard for portfolio and trade-level analysis
- > Driven by Bloomberg market data

Credit Value Adjustment (CVA) | Initial Margin (IM) | Derivatives pricing

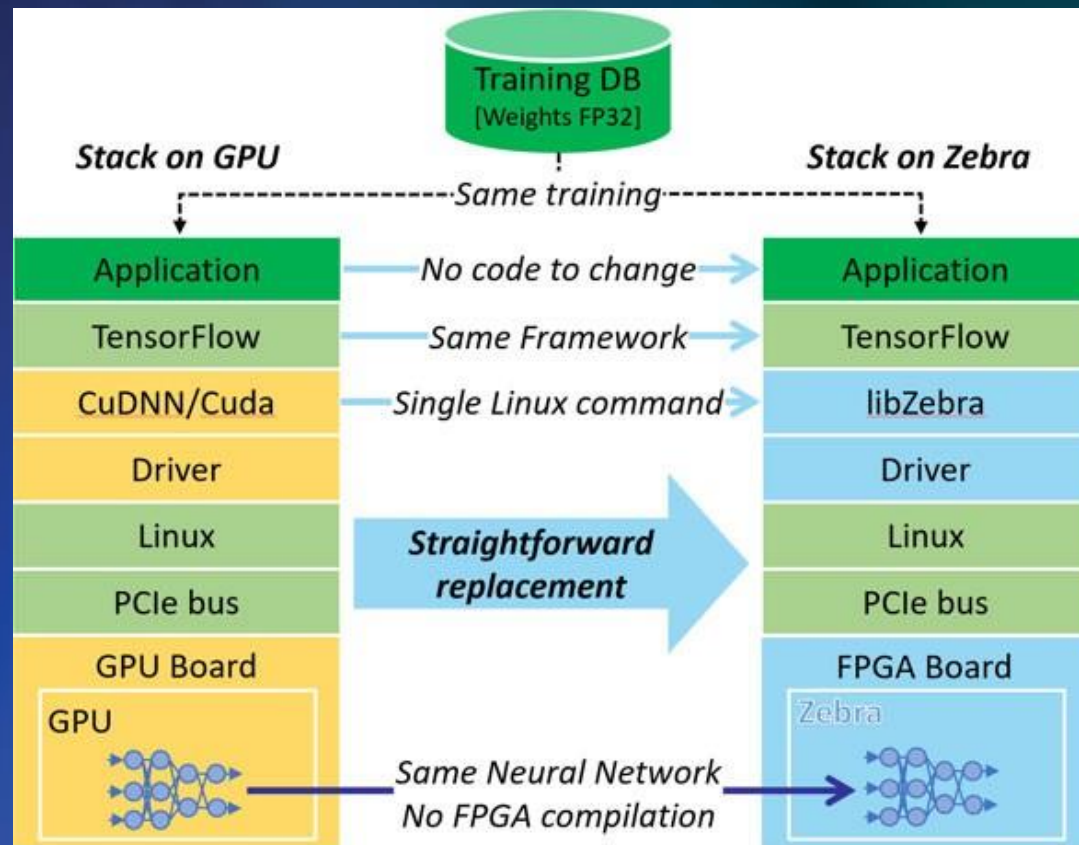
Accelerated Genomics Pipelines



- > No proprietary pipelines.
- > 10x Speed Up over Original GATK pipelines
- > Supports multiple GATK versions for germline and somatic pipelines

Genome analysis on industry standard GATK pipelines | No FPGA Experience required

Neural Network Inference



- > Uses parameters from GPU training
- > No proprietary training or re-training needed
- > Offers same accuracy as GPU or CPU
- > Fully off-load CPU from all neural network computing
- > Up to 3 billion weights in a single network
- > Up to 1 million layers

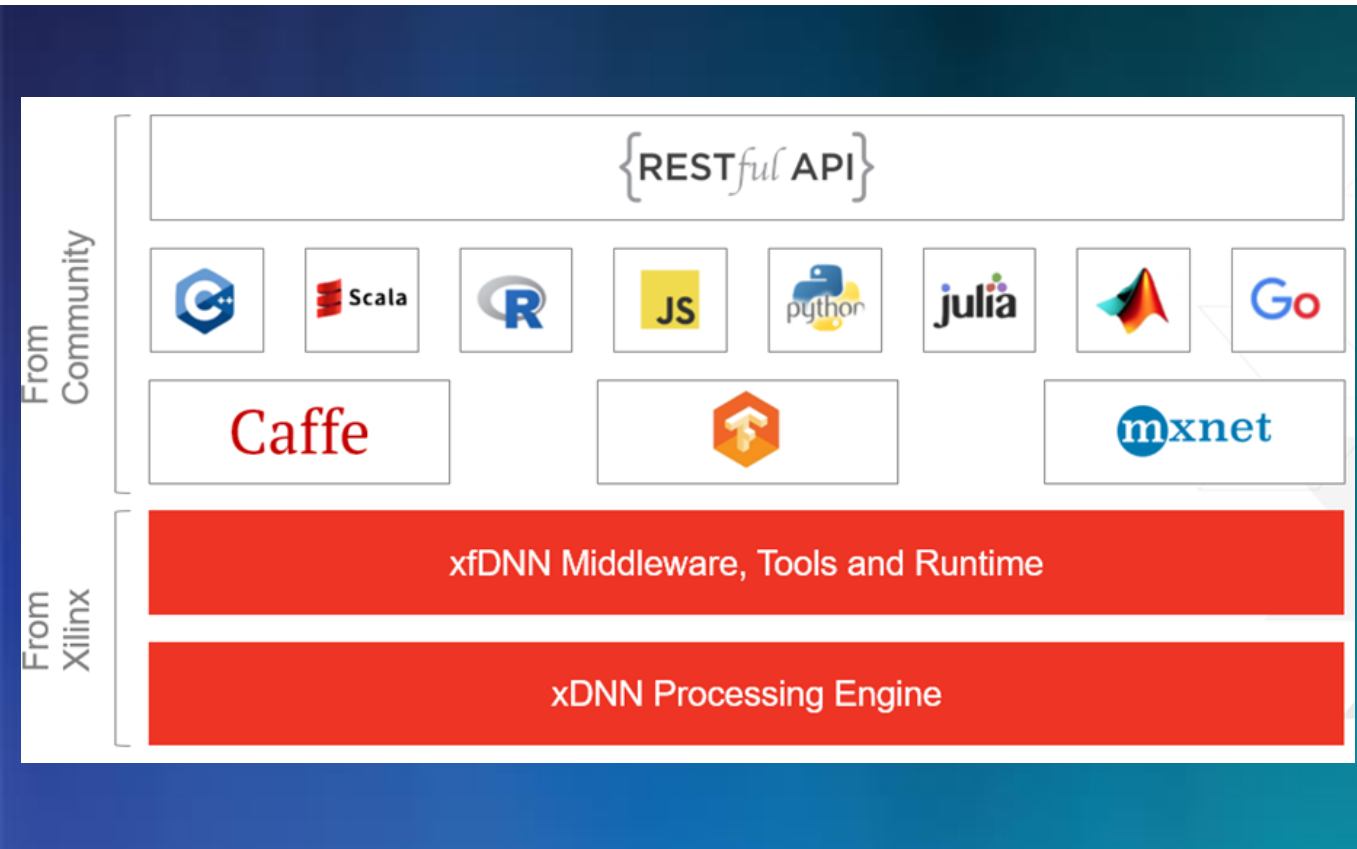
AlexNet | CaffeNet | GoogLeNet V1 | Inception V3/V4 | VGG16/19 | ResNet50/152

Real-Time Text Analytics



- > Up to 100x speed-up for text analytics
- > Topic identification & summarization
- > Sentiment & consensus analysis
- > Python/ Accelerator hybrid library - core analytics on Alveo and distributed-CPU for peripheral analytics.

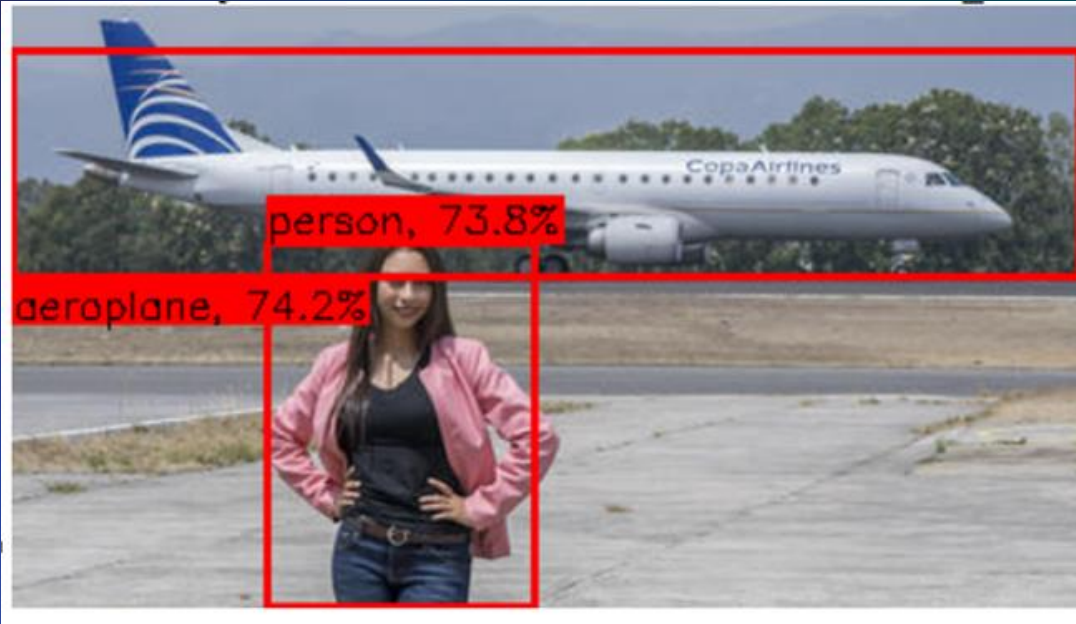
Financial Services | Government Intelligence | Social Media Analytics



- > **xDNN IP** - High Performance general CNN processing engine.
- > **xDNN Middleware** - Software Library and Tools to Interface with ML Frameworks and optimize them for Real-time Inference.
- > **ML Framework and Open Source Support** - Support for high level ML Frameworks and other open source projects.

High throughput low latency inference: Caffe | Tensorflow | Mxnet

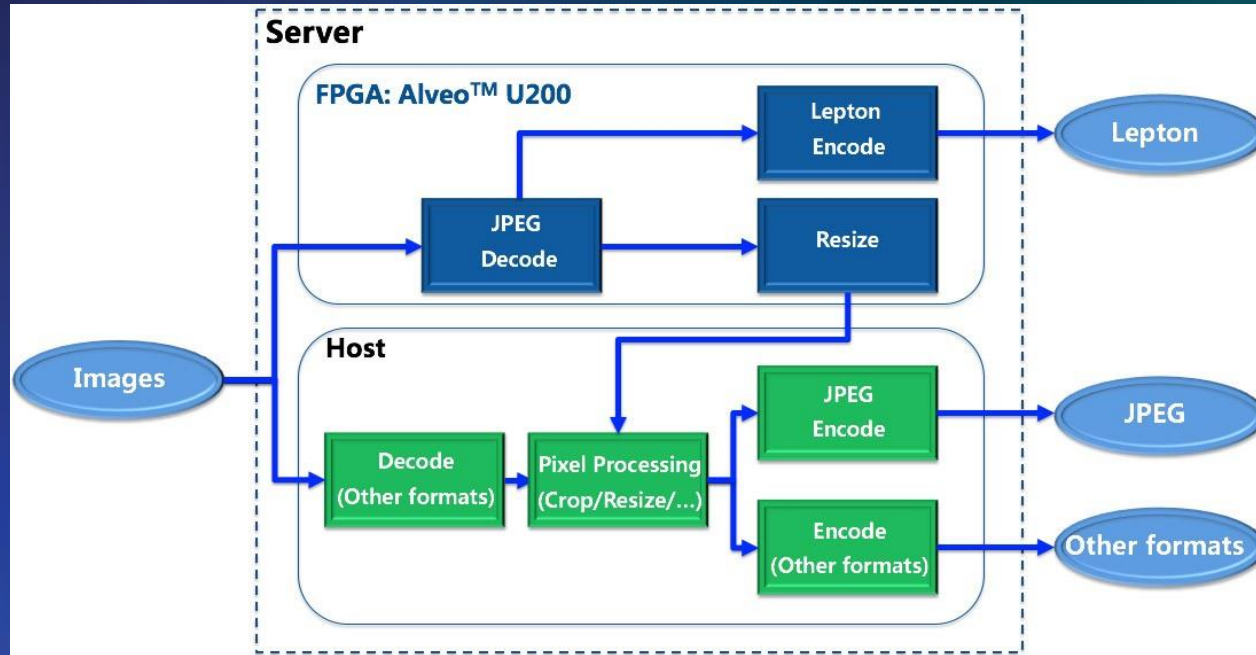
Image and Video Edge Analytics Acceleration



- > Maximizes performance of FPGA technology versus doing image/video machine learning
- > Uses GPU or CPU trained Convolutional Neural Networks on FPGAs doing inference analysis
- > Achieves reliable, accurate results with smaller, low-power solution

Object detection and Categorization on video | Fine grained classification

Image Processing Accelerator



- > Improves the performance of image processing and image analytics
- > Throughput Gain: Up to 6 times
- > Latency Reduction: Up to 4 times
- > Reduce TCO by up to 3x

ImageMagick | OpenCV | GraphicsMagick | Lepton

Personalized Streaming Video



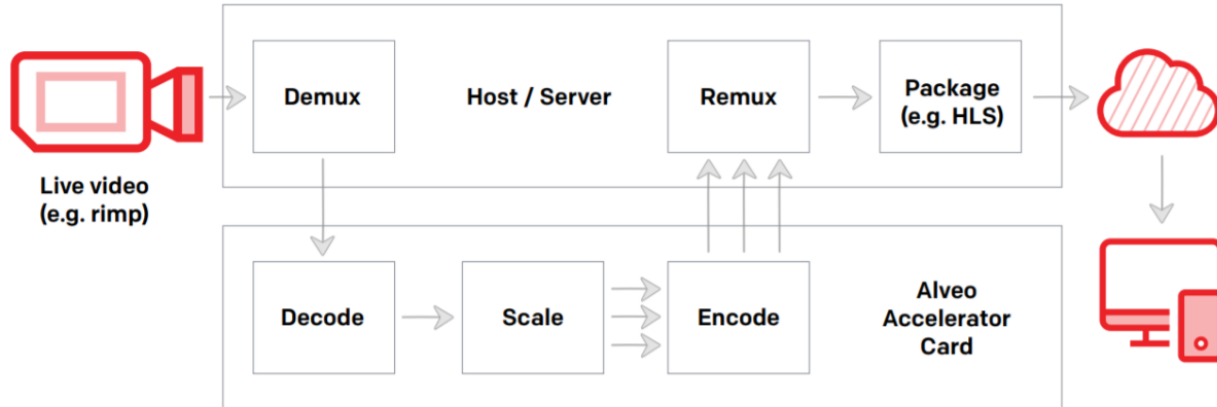
- > Real time encoding of 4 HD Streams
- > Video + ML Solution
- > 3-Layer IML
- > Unlimited User Layouts
- > Skreens Launchpad application manages and reconfigures Cloud configuration

Customer Tailored Live video Layout | Video Intelligence

Adaptive Bit Rate Video Transcoding



FFMPEG TRANSCODING APPLICATION



- > High performance HEVC & VP9 encoders
- > Easy programming with FFmpeg integration
- > Fully configured transcoding pipeline
- > 32 simultaneously independent encoded streams on a single Alveo

FFmpeg plugins for Decoder, Scaler, and Encoders | Download Free Trial on Alveo



FAST

Faster than CPUs & GPUs

Latency advantage over GPUs



ADAPTABLE

Optimized for any workload

Adapt to changing algorithms



ACCESSIBLE

Deploy in the cloud or on-premises

Applications available now



Full partner details:

<https://www.xilinx.com/products/boards-and-kits/alveo/applications.html>

© Copyright 2018 Xilinx

