

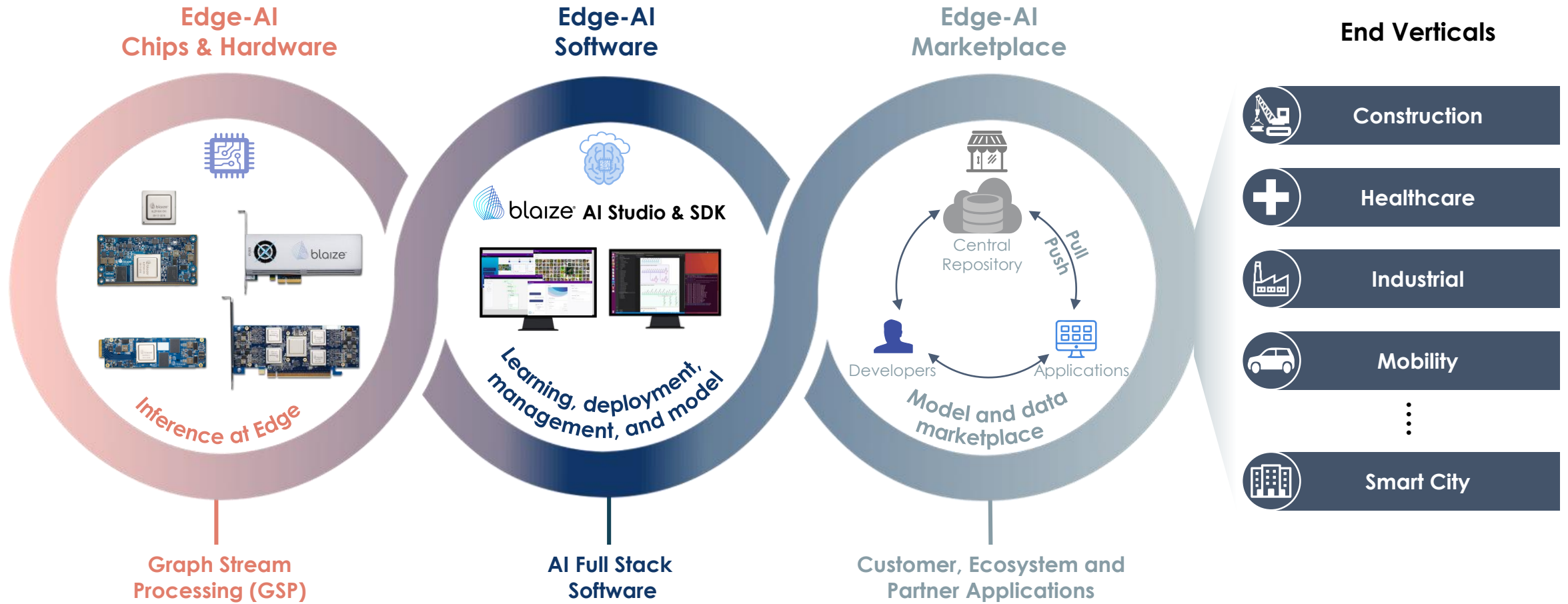


Developing Sustainable Edge AI Products for the Real World

Barrie Mullins
Head of Marketing
Blaize

Blaize is a Full-Stack Solution Enabling Edge AI

Integrated-platform increases ecosystem value over time



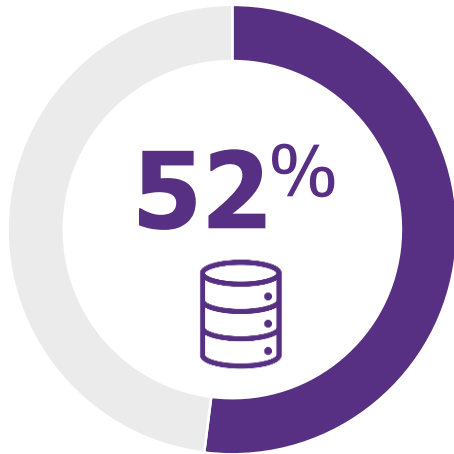
AI Modeling/Software Expertise and Time Spent Are Key Impediments to Adoption



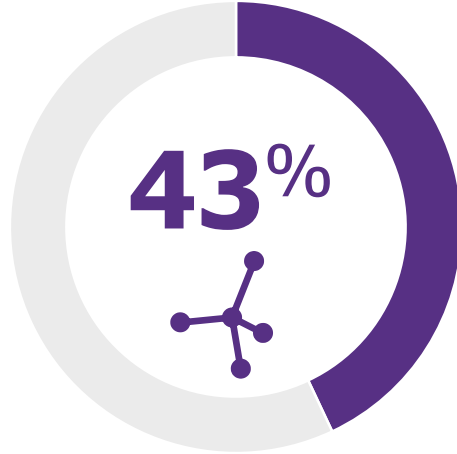
“87% of data science projects never make it into production”

—Chris Chapo, SVP of data and analytics at Gap¹

Top Factors Holding Back AI Deployments²



**Data Quality,
Quantity, Access**



**Explainability of
Algorithms and Selection**



**Skilled Data
Science Personnel**



Access to Edge-specific application data



Inefficiency of x86 and GPU trained models



Time required to train and optimize models with accuracy and performance for the Edge



Large annotated and labeled data sets



Limiting parameters of existing software solutions (power, performance, latency)

¹ "Transform 2019, San Francisco, Venture Beat article

² "Market Analysis Perspective: Worldwide Artificial Intelligence Software, 2020", IDC

Enabling the Edge to Become the Physical-World-Equivalent of the Internet...

Poised to capture edge **data** and deliver **insights**



Democratizing Knowledge

Democratizing Data

KNOWLEDGE

DATA



MARKET PLACE

Develop AI

AI Knowledge Platform

AI Studio

Deploy and Manage AI

AI-Created AI



INDUSTRIAL



SMART CITIES



MOBILITY



AGRICULTURE

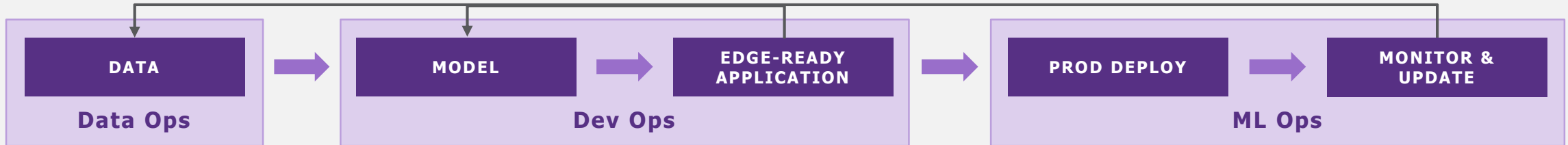
Days to Development

Days to Deploy

Blaize AI Software Suite



AI Studio



Open Tools and ML Frameworks

Picasso SDK

Dev Kit <ul style="list-style-type: none">Integrated Dev EnvironmentDebuggerPerformance Profiler	GSF Demo Suite NetDeploy	Graph Framework (OpenVX, OpenCL, C++) <ul style="list-style-type: none">Graph CompilerGraph RuntimeCustom Kernel Compiler
---	---	--

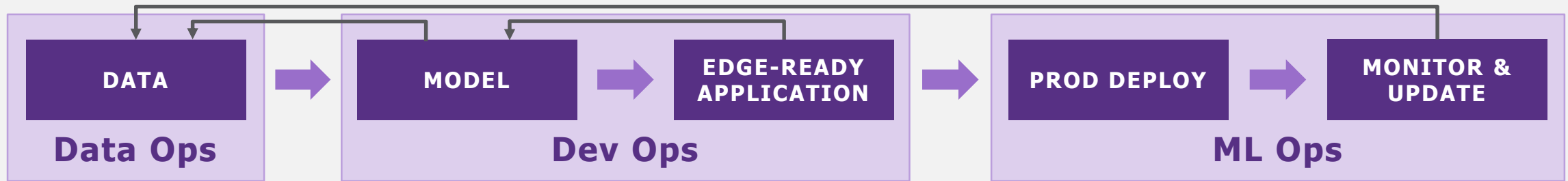
Graph Libraries + AI Toolkit (OpenVX, GStreamer, Python)

Neural Network Library	Computer Vision Library	Automotive Library	Image Signal Processing Library	Sensor Fusion Library	FFT Library
------------------------	-------------------------	--------------------	---------------------------------	-----------------------	-------------

Enabling AI from Concept to Deployment Over the Product Lifecycle



AI Studio



- Collect
- Explore
- Label & annotate
- Share via Marketplace

- Import or Train (TL)
- Validation and Testing
- Pre & Post Processing
- Versioning
- Share via Marketplace

- Quantize
- Fine Tune
- Build application

- Deploy to client-side device
- Configure sensors
- Obtain real-time updates
- Obtain resource usage and operating metrics

- Monitor model for drift (model degradation)
- Review snapshot images from streaming input source
- Label/annotate and Retrain model

Assistive Help

Pathfinder GSP Embedded Platforms



Smart Vision Use Cases

- Traffic & Parking Management
- Retail Camera Systems
- Autonomous Optical Inspection
- Robotics
- Embedded PCs
- Network Edge Equipment

PATHFINDER P1600 INDUSTRIAL EMBEDDED SYSTEM ON MODULE



ARM Cortex A53, 4GB LPDDR4,
Ethernet, CSI, USB, CODEC, I/O...
Industrial Grade

16
GSP CORES

16
TOPS

10 W
AVERAGE POWER

**100% SW & HW PROGRAMMABLE
DESIGNED FOR PERFORMANCE / WATT**



Blaize Xplorer Accelerator Platforms



Inference AI Use Cases

- Edge and Enterprise Servers and Appliances
- Autonomous Optical Inspection
- Smart Parking and Traffic Management
- Smart Retail Applications
- Network Video Recorders
- Warehouse and Factory Applications
- Safety and Security Systems
- Industrial PCs & Servers



XPLORER™ ACCELERATOR PLATFORMS

Xplorer X1600E EDSFF



Xplorer X1600P PCIe

PCIe 3.0 x4, 4GB LPDDR4
EDSFF E1.S
Enterprise Grade

16 GSP CORES

16-20 TOPS

10 W

AVERAGE POWER

Xplorer X1600P-Q



Xplorer X1600P-Q



PCIe 3.0 x4, x16, 16GB LPDDR4
Commercial & Enterprise Versions

64 GSP CORES

80 TOPS

65 – 80 W

AVERAGE POWER

**100% SW & HW PROGRAMMABLE
DESIGNED FOR PERFORMANCE / WATT**

Blaize Edge Inference Deployment

Cities, Factories, Retail, Safety, Security, Agriculture and More...



AI Studio Host Edge Inference

AI Studio Client/Picasso Edge Inference

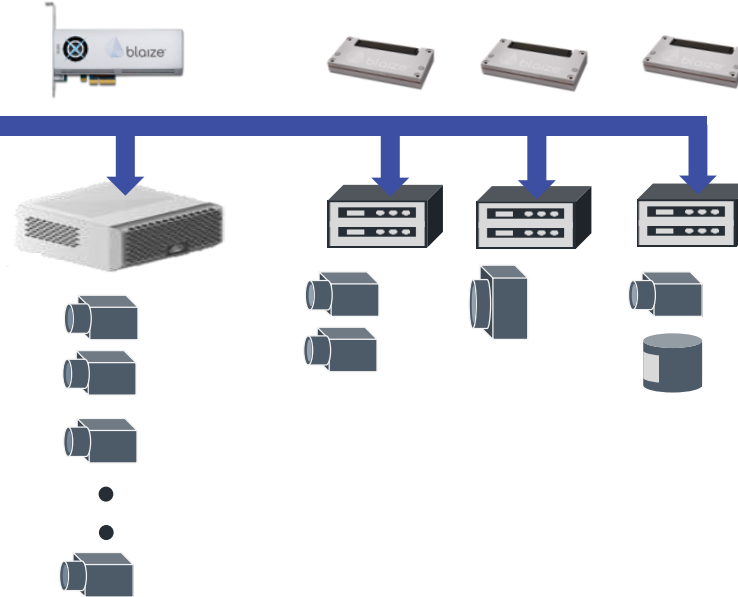
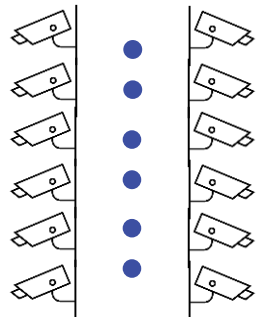
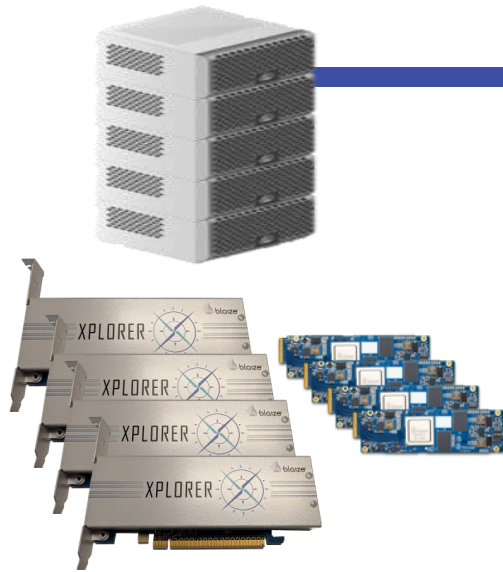
Edge Applications

AI Development & MLOps Server/Cloud

AI Development & ML Ops Edge Server

Industrial PC

Embedded SOM



Software Complete Development Suite

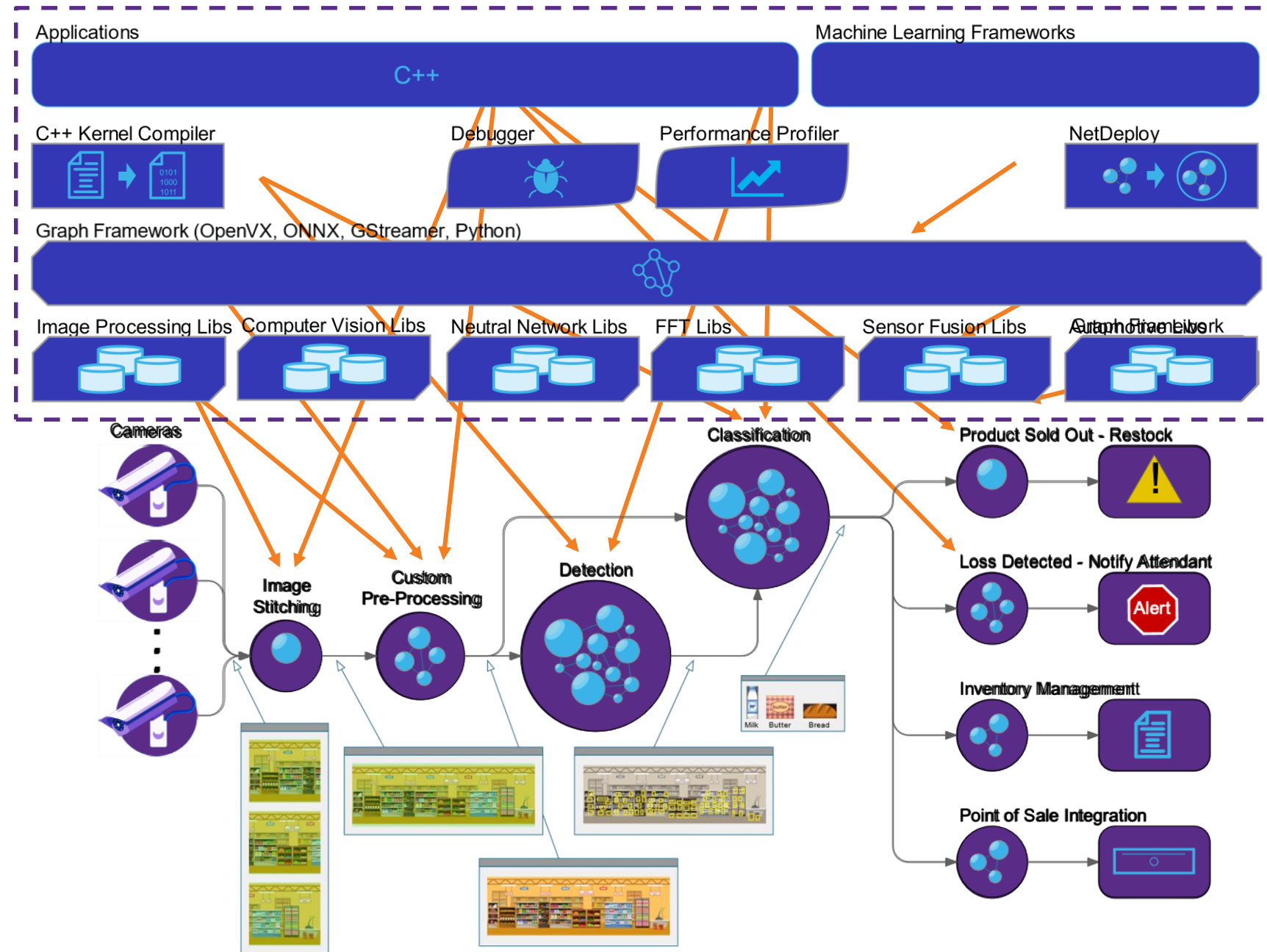
Software Dev Kit (SDK)

- Complete tools, libs, etc.
- Vision, NLP, RM, Fintech
- Foundation of AI Studio

End-to-End Application Development

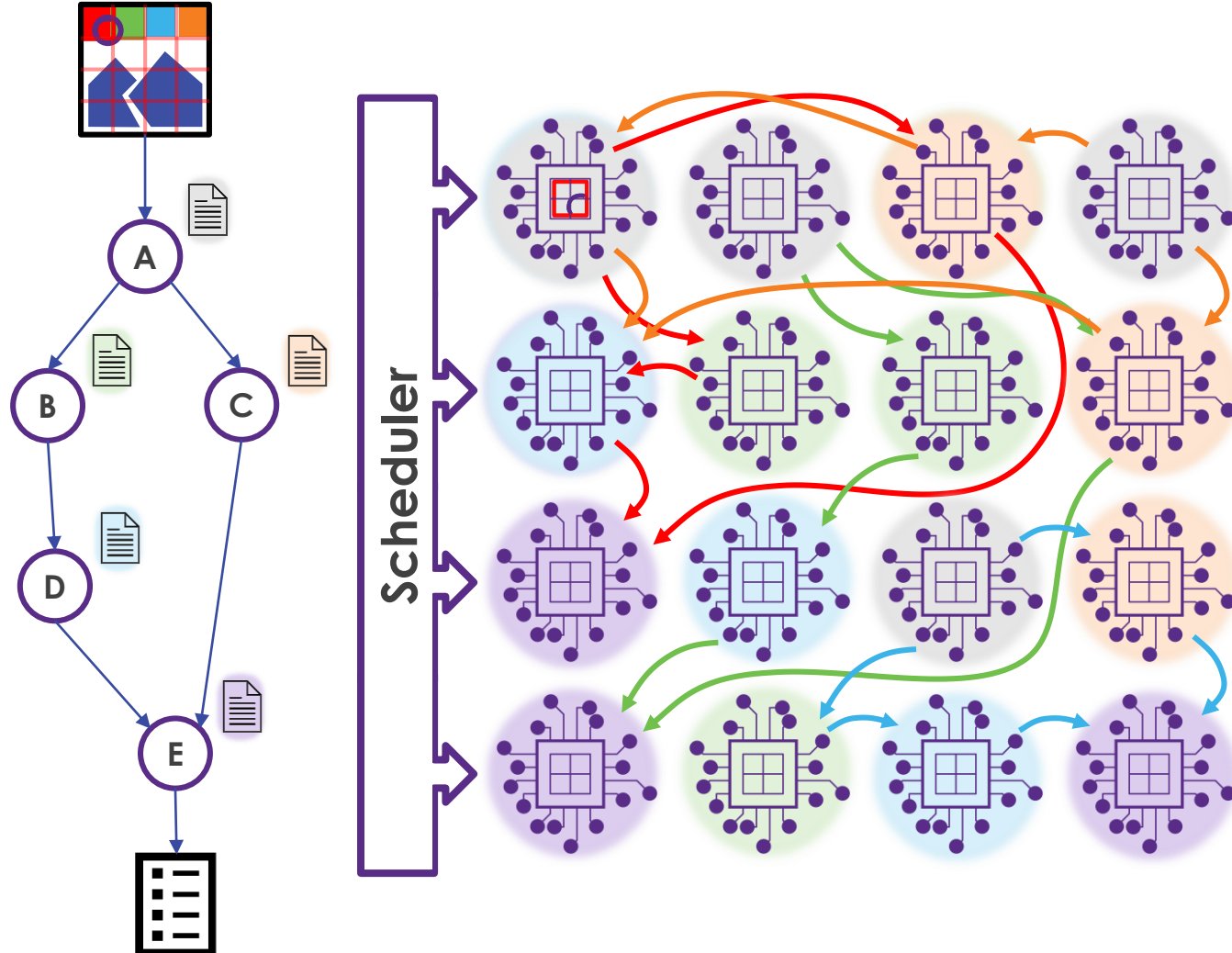
- Convert/Optimize ML Models
- Custom C++ Programs

Example: Grocery Store



Executing Programs in a Data Flow Manner

Graph Streaming Processor (GSP) Architecture



Graph Streaming Processor

- ✓ Increases hardware utilization without batching
- ✓ Reduces latency with on-chip data movement
- ✓ Reduces power consumption
- ✓ Less “dark silicon” or fallow hardware
- ✓ Dynamically allocates compute resources for rapid graph changes
- ✓ Supports general purpose data parallel program execution, not just AI

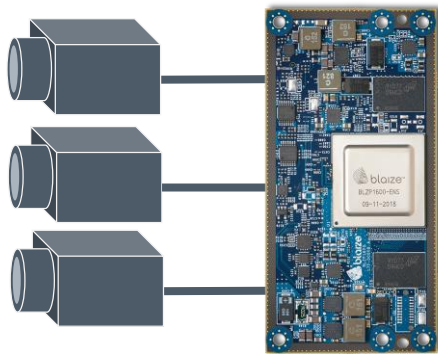
Case Study Retail: Edge AI Retail Monitoring Systems for Customer Safety and Service

Blaize AI Accelerators enable customer safety and service with low power, low latency multicamera system

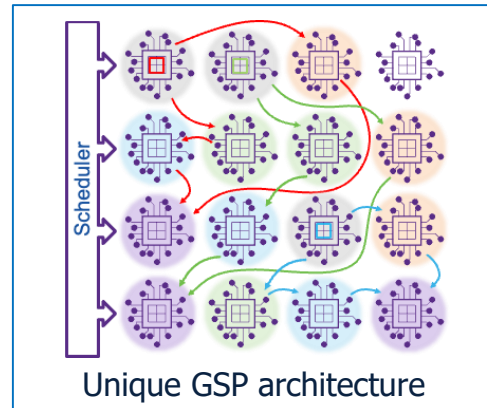


Blaize AI Accelerators delivers power efficient multicamera system with alerts and warnings

- ① **Multi Camera Edge System**
Detecting customers, hazards and products

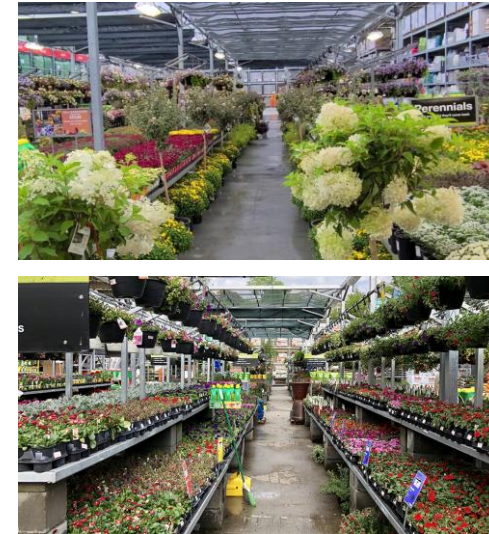


- ② **AI Model Execution**
Processing multiple AI models on live streaming data



Unique sensors processed and merged efficiently

- ③ **Application Detection**
Safety and detection provided with performant AI models running at the edge



- ✓ Support for input streams from many different sensors

- ✓ 2x the # of sensors and performance over competitor solution

- ✓ Integrated high-performance solution

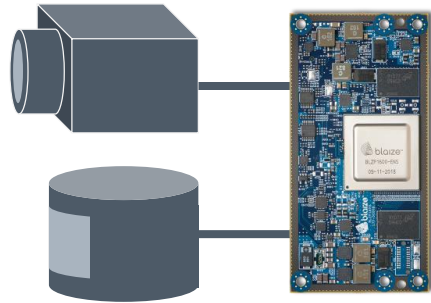
Case Study Mobility: High Performance Sensor Fusion

Accelerating high performance in sync Camera and Lidar processing in a low power, low latency system

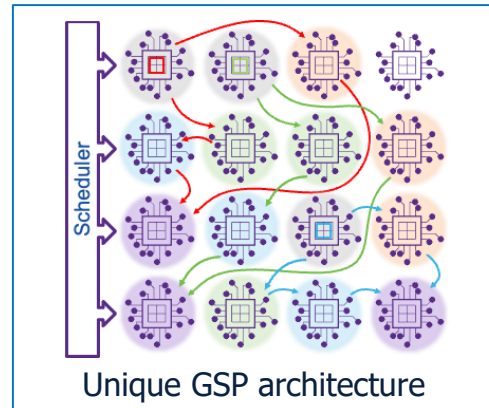


Blaize AI delivers **2x performance** at the same power over competitor system

- ① **High Speed Sensor Fusion**
Capturing and merging video and point cloud data

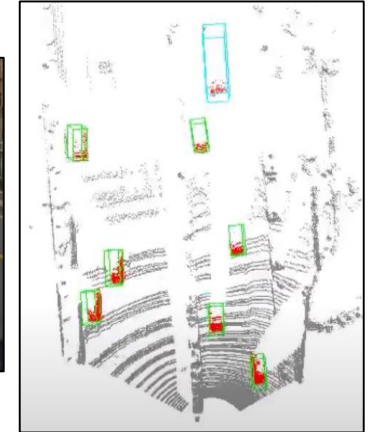


- ② **AI Model Execution**
Processing multiple AI models on live streaming data



Unique sensors processed and merged efficiently

- ③ **Application Detection**
Insight and visibility provided with performant AI models running at the edge



- ✓ Support for input streams from many different sensors

- ✓ 2x the # of sensors and performance over competitor solution

- ✓ Integrated high-performance solution

Adaptable Edge Systems Require



- 1** Full Stack End-to-end AI Software Suite
- 2** New Processing Architecture for Low Latency and Low Power
- 3** Ability to Sustain Products over Their Lifecycle

Thank You



**Visit us at
Booth 506**



**Open for
Customer
Meetings**



Demos with our Partners

- Embedded Facial Recognition
- AI Studio on Blaize Infrastructure

• INNOVATRICS
building a world of instant trust

**Looking for Partner
Engagements**



Thank You

