

# Introducing the Vitis Unified Software Platform

隆重介绍 Vitis 统一软件平台

**Salil Raje**

**Executive Vice President 执行副总裁**

**& GM Data Center Group 数据中心事业部总经理**





# WELCOME

All Developers

拥抱所有开发者



**Heterogeneous  
Compute**  
异构计算



**Cloud to Edge**  
从云到边缘



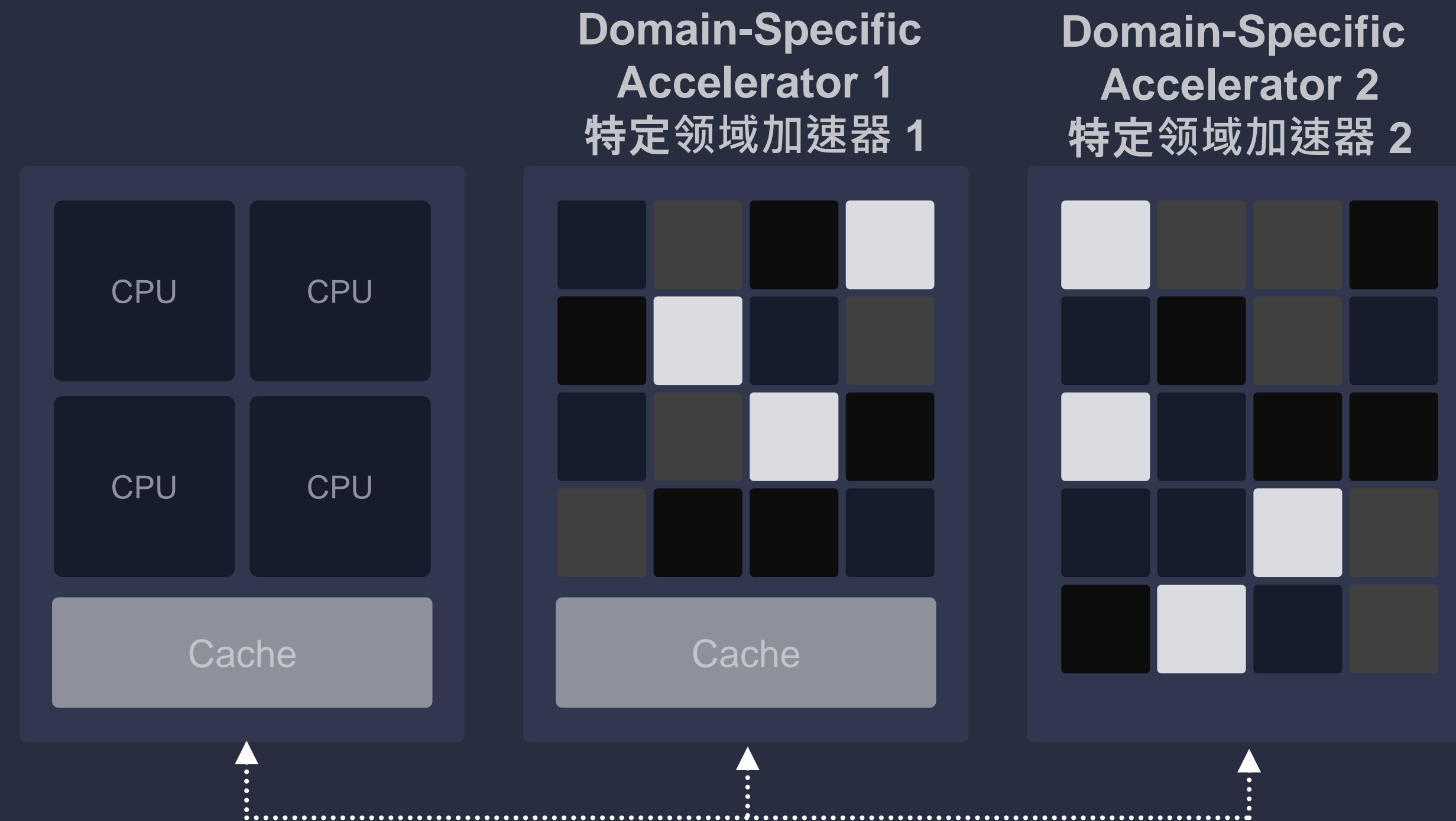
**AI Proliferation**  
AI 激增

# Industry Trend: Heterogeneous Compute

## 行业趋势：异构计算



Engines Customized to Accelerate Specific Domains  
定制引擎以加速特定领域应用



### Key Challenge 主要挑战

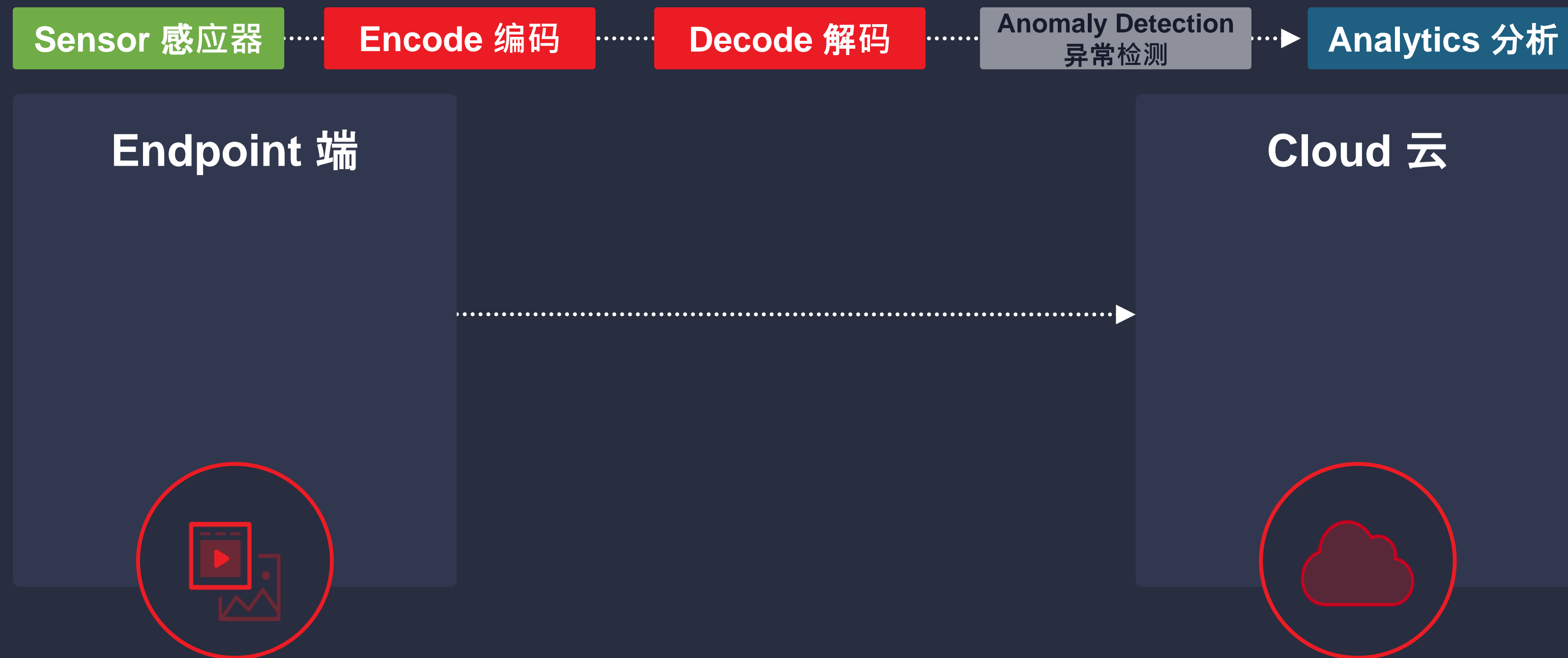
Programming & Integration of Adaptive Acceleration Engines  
自适应加速引擎的编程与集成

# Industry Trend: Cloud to Edge

## 行业趋势：从云到边缘



Applications are often split between cloud and edge  
应用总是分隔在云和边缘



# Industry Trend: Cloud to Edge

## 行业趋势：从云到边缘



Applications are often split  
between cloud and edge  
应用总是分隔在云和边缘

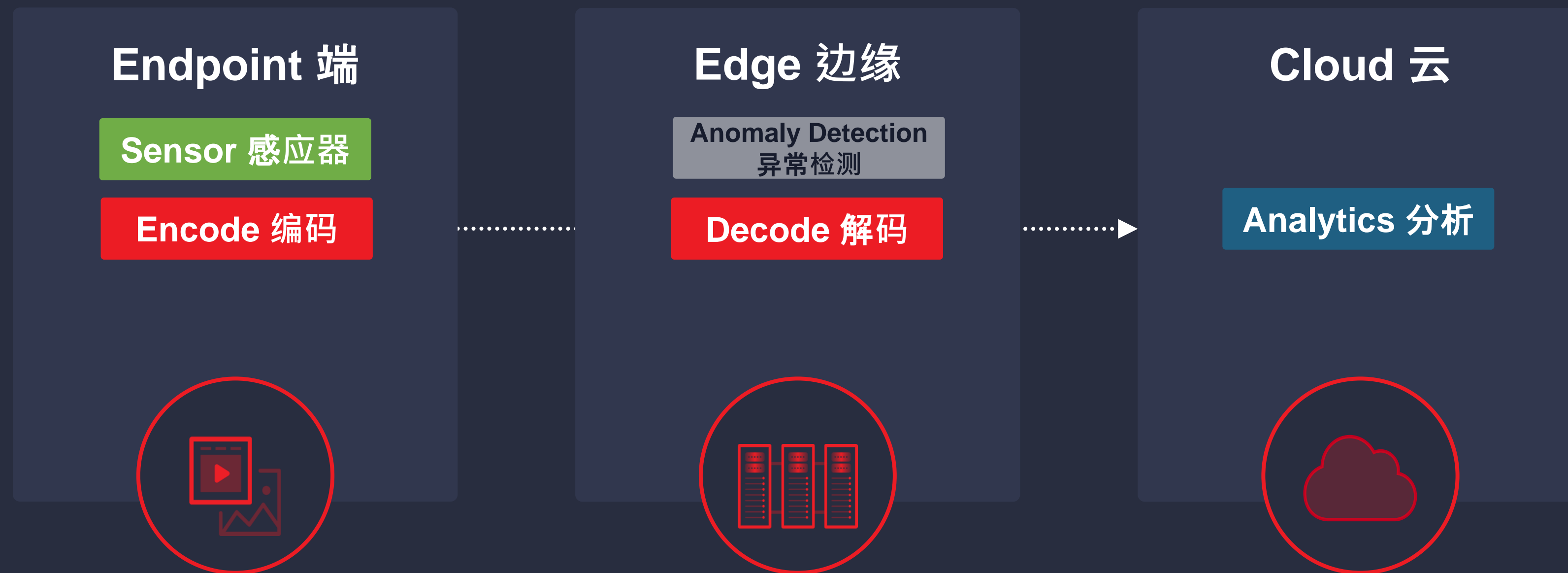


# Industry Trend: Cloud to Edge

## 行业趋势：从云到边缘



Applications are often split  
between cloud and edge  
应用总是分隔在云和边缘

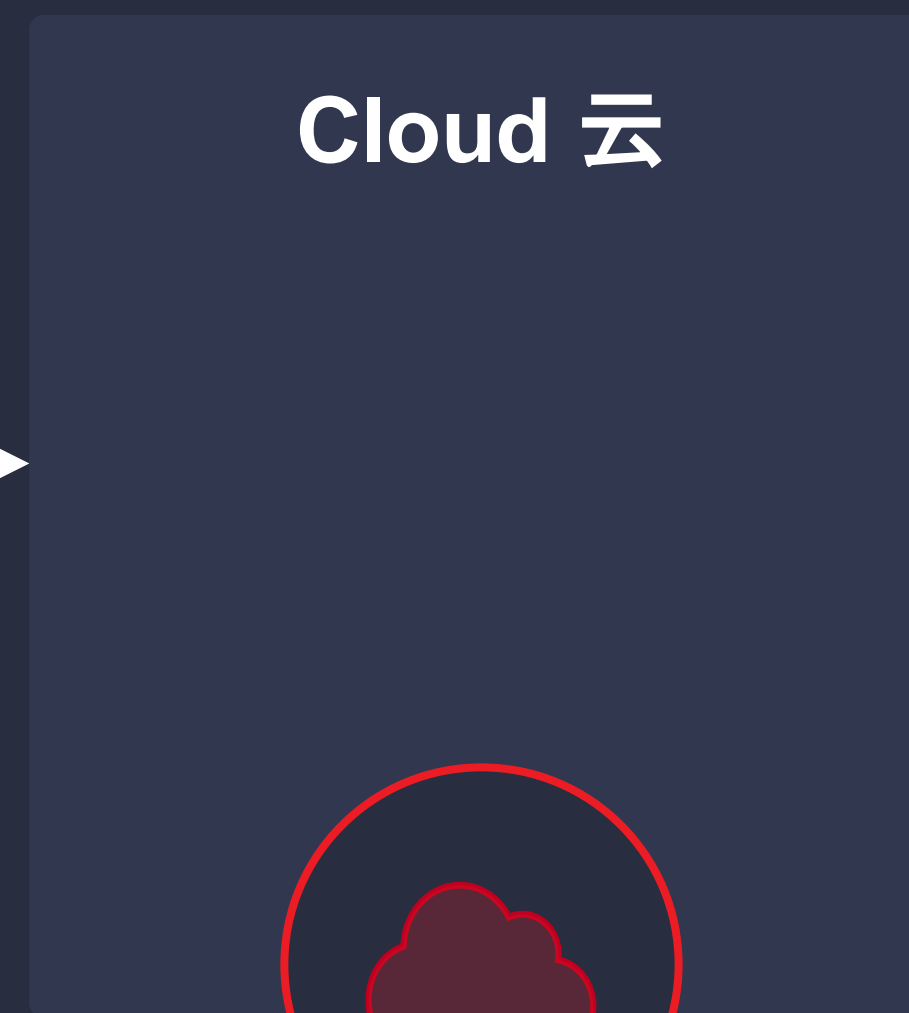
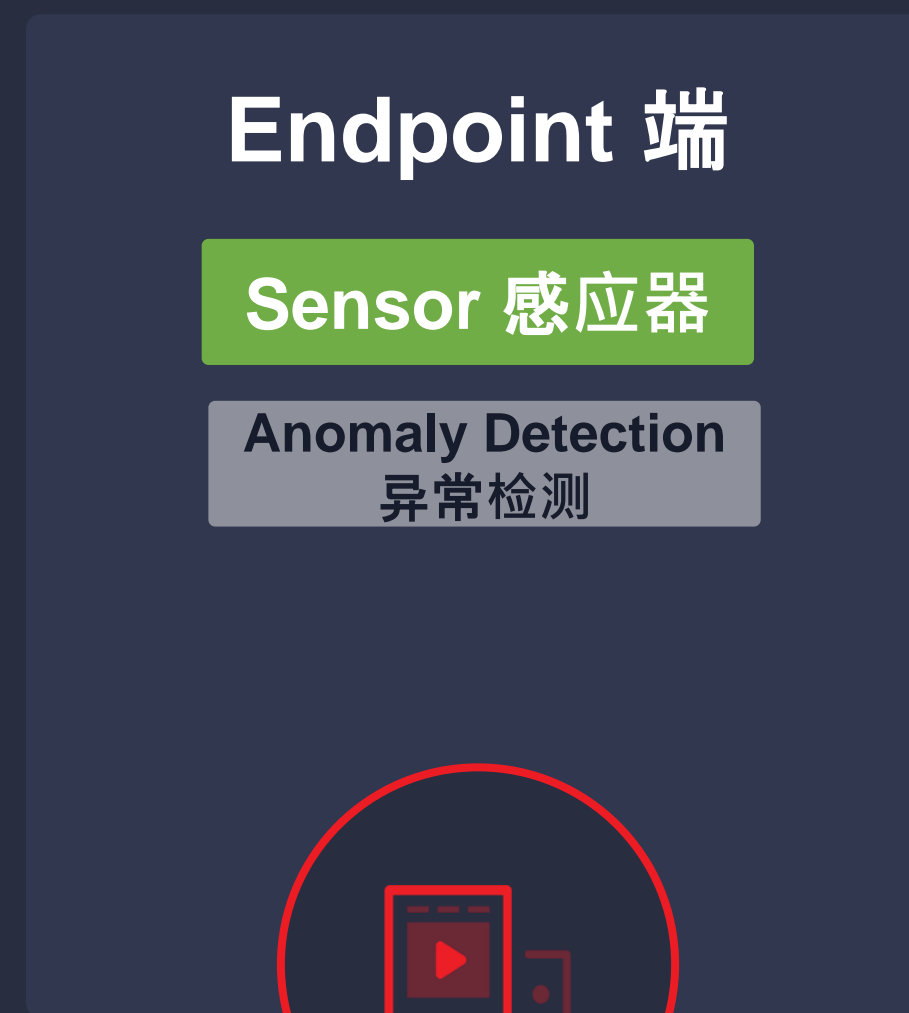


# Industry Trend: Cloud to Edge

## 行业趋势：从云到边缘



Applications are often split between cloud and edge  
应用总是分隔在云和边缘



**Key Challenge**  
主要挑战

Need for Retargetability  
需要可重定目标的能力

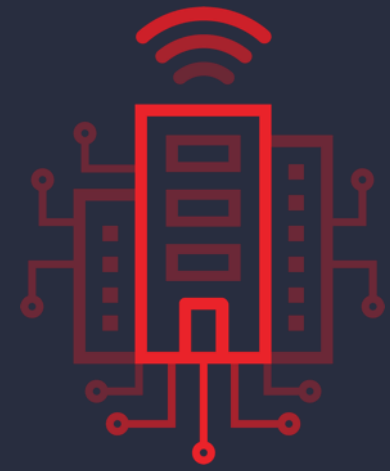


# Industry Trend: AI Proliferation

## 行业趋势：AI 激增



AI is being used in many applications  
AI 正被用于众多应用



Smart City  
智慧城市



Smart Retail  
智慧零售



Autonomous Driving  
自动驾驶



Security  
安全



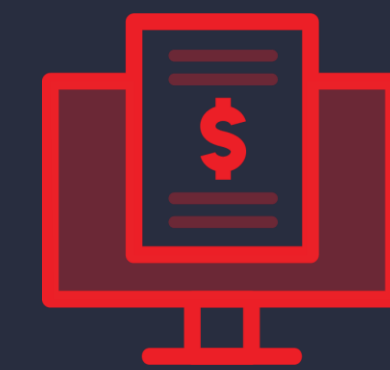
Genomics  
基因组分析



Video Analytics  
视频分析



Healthcare  
医疗



Finance  
金融

## Key Challenge 主要挑战

Acceleration and  
Integration of the  
Whole Application  
整体应用的加速与集成



**Heterogeneous  
Compute**  
异构计算



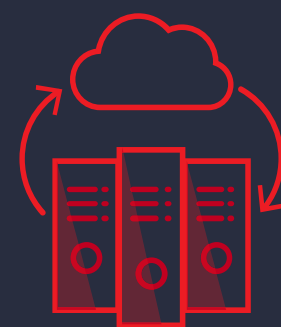
**Cloud to Edge**  
从云到边缘



**AI Proliferation**  
AI 激增



**Heterogeneous  
Compute**  
异构计算



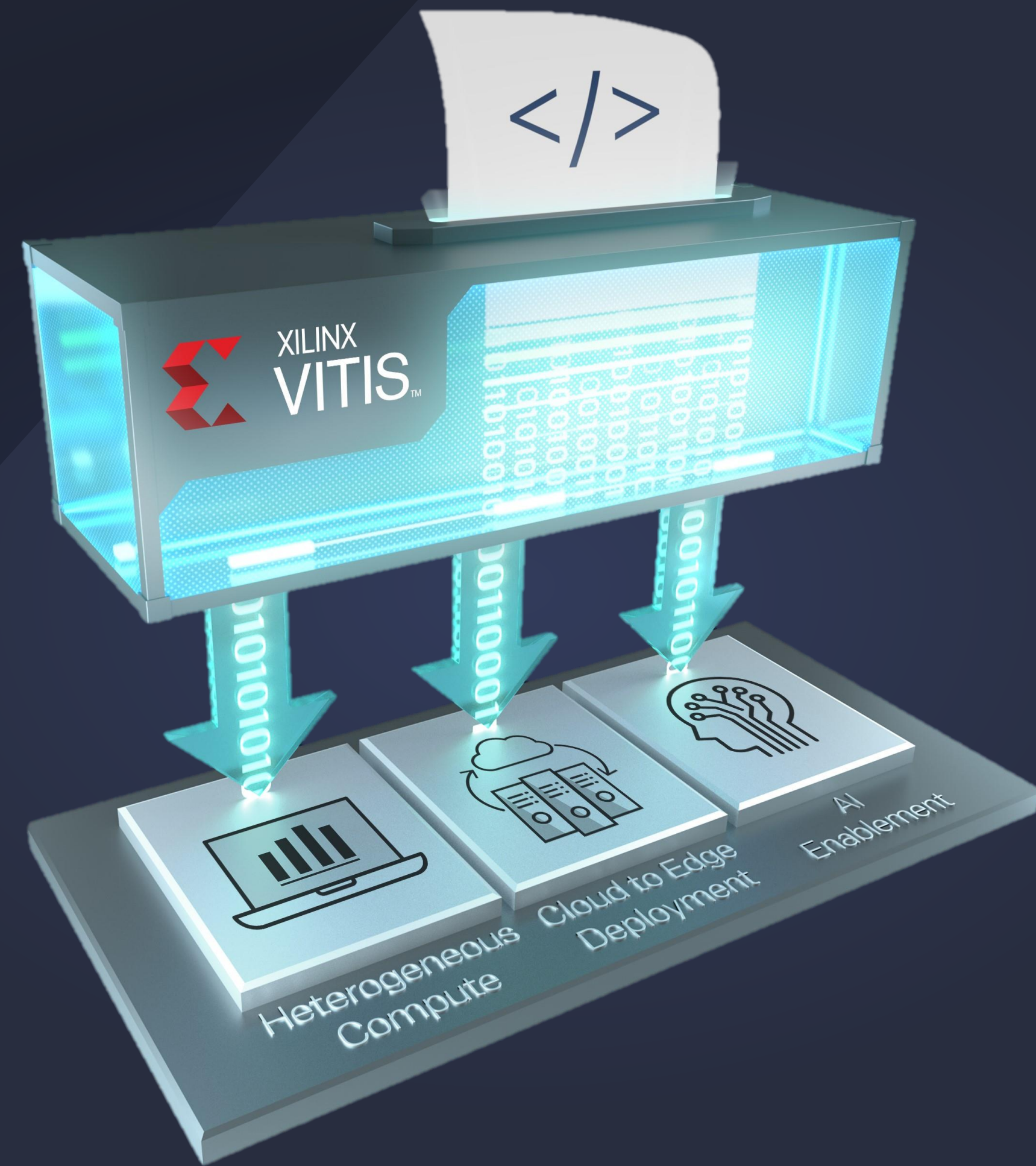
**Cloud to Edge**  
从云到边缘



**AI Proliferation**  
AI 激增

# Vitis Unified Software Platform Vitis 统一软件平台

---

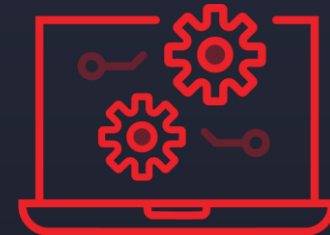


# Platform Transformation

## 平台转型之路

#DEVELOPERS

#开发者

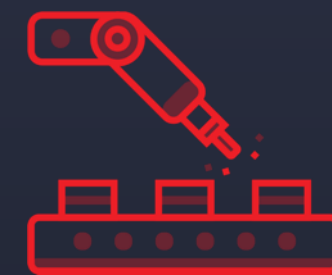


Vivado

2012



OS and  
Firmware SDK  
操作系统和固件SDK



SDSoC, Embedded  
SDSoC、嵌入式



SDAccel, Data Center  
(FaaS, Alveo)  
SDAccel、数据中心  
(FaaS、Alveo)



AI inference  
Acceleration  
AI 推断加速



Vivado

2019

Vitis Unified  
Software Platform  
Vitis统一软件平台  
Adaptable & Programmable  
灵活应变 & 可编程



# Enables All Developers to Build and Deploy to All Platforms

支持所有开发者构建和部署各种平台



Build  
构建



Embedded  
Developers  
嵌入式开发者



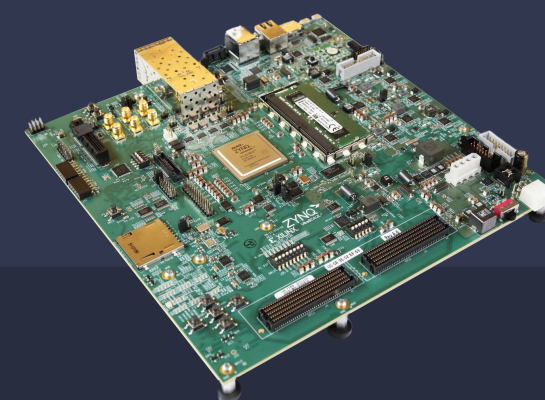
Enterprise  
Application Developers  
企业应用开发者



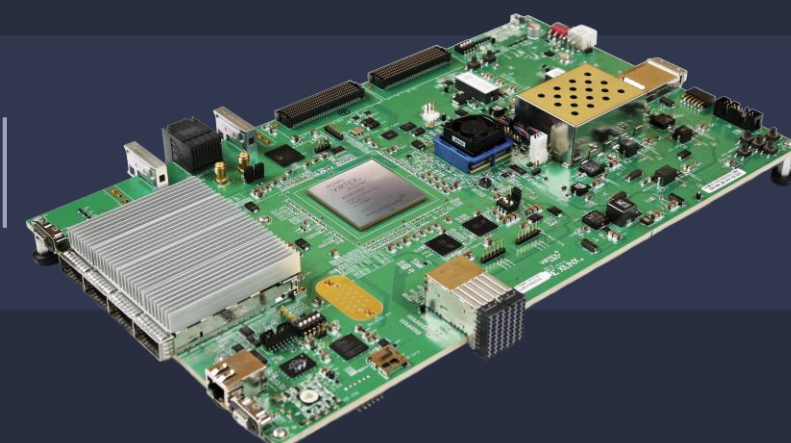
Enterprise Infrastructure  
Developers  
企业基础设施开发者



Data & AI  
Scientists  
数据和 AI 科学家



Zynq



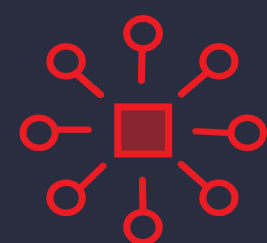
UltraScale



Alveo

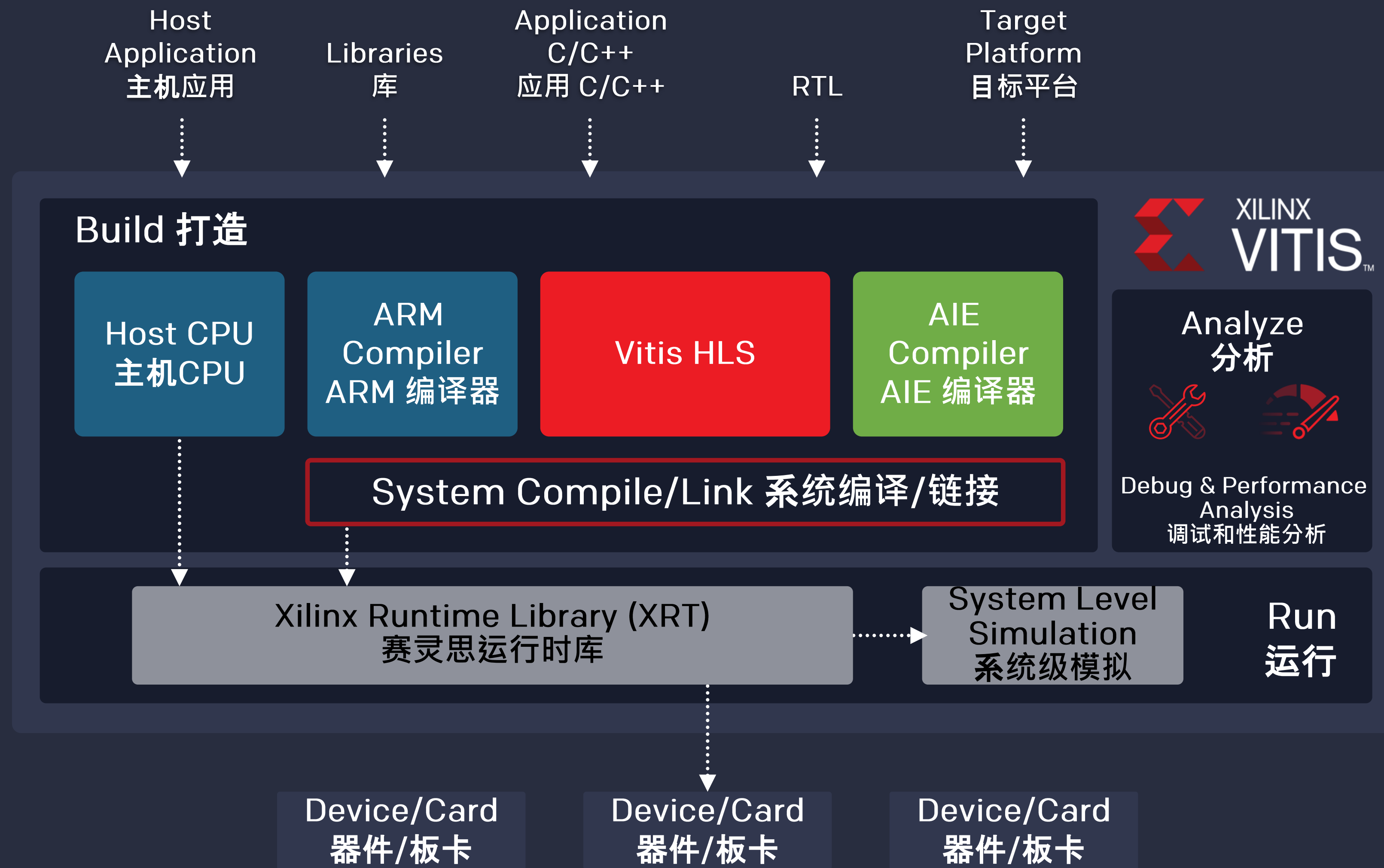


Data Center Rack  
数据中心机架



Deploy  
部署

**Build**  
**Comprehensive Development Tool Suite**  
打造丰富的开发工具套件



400+ functions across multiple libraries 多种库提供 400 多项功能

Open-Source, performance-optimized out-of-the-box acceleration 开源、性能优化的开箱即用加速

Build

# Comprehensive Development Tool Suite

打造丰富的开发工具套件

## Domain-Specific Libraries 特定领域库



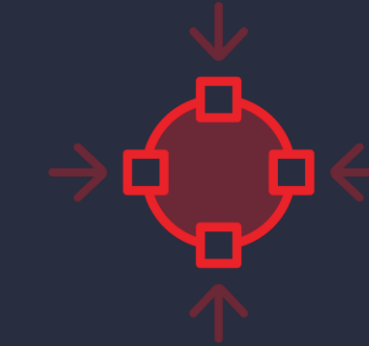
Vision &  
Image  
视觉与图像



Finance  
金融



Data Analytics &  
Database  
数据分析和数据库

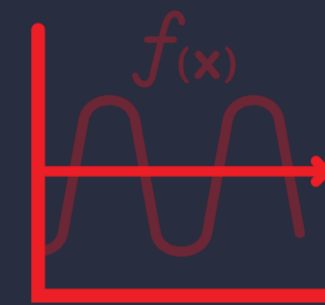


Data Compression  
数据压缩

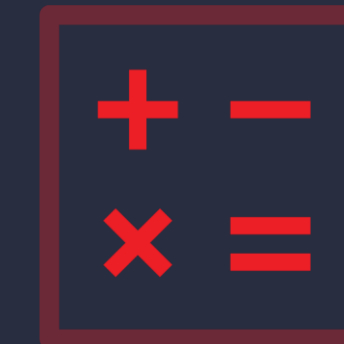


Data Security  
数据安全性

## Common Libraries 通用库



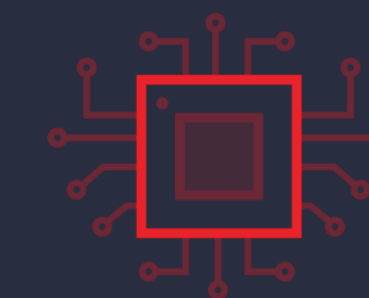
Math  
数学



Linear Algebra  
线性代数



Statistics  
统计



DSP  
数字信号处理器



Data Management  
数据管理

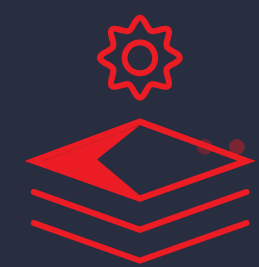




Deploy

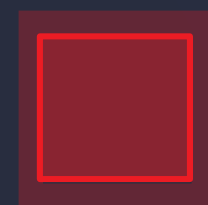
# Embedded Deployment 嵌入式部署

Application  
应用

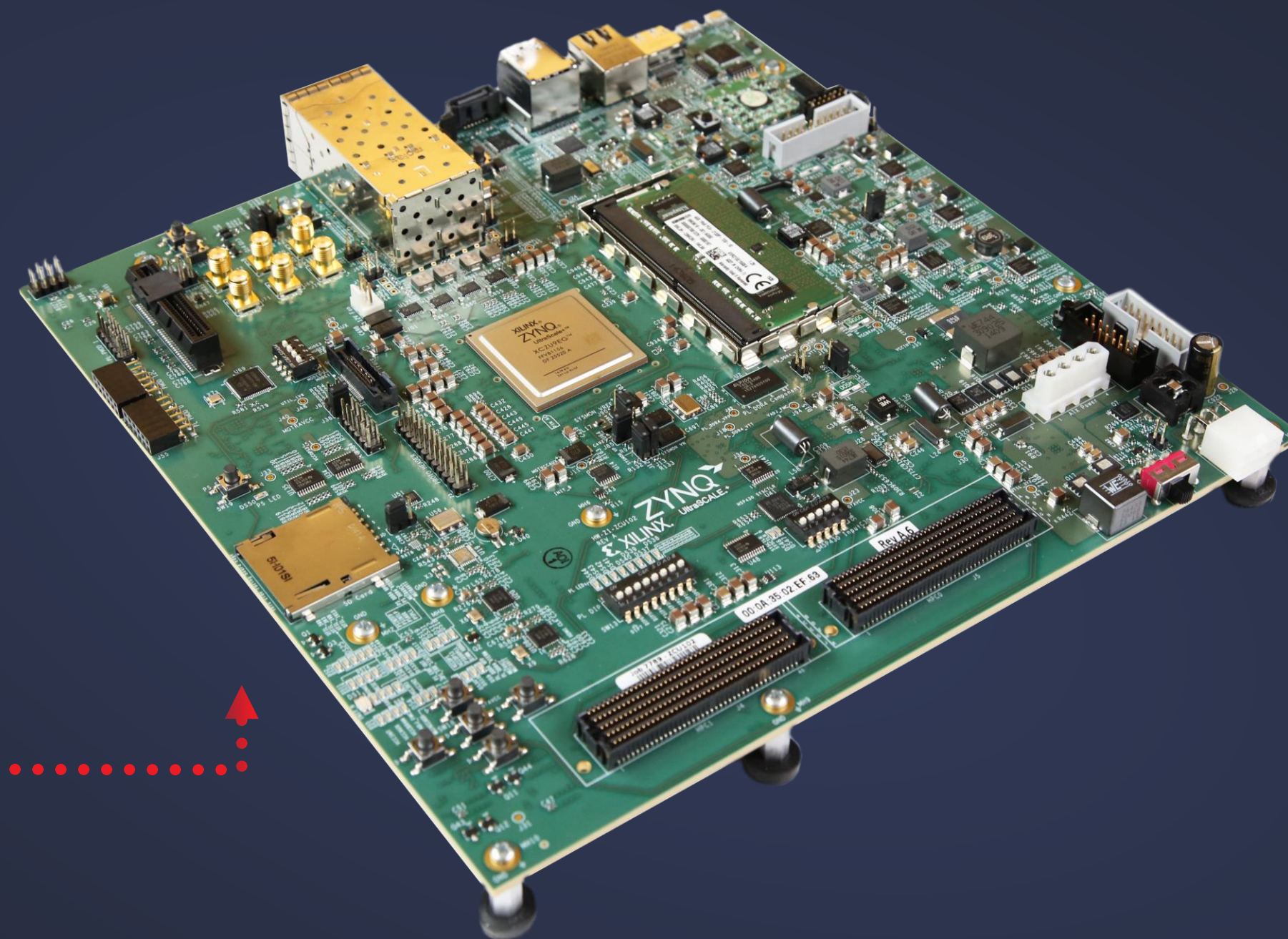


XILINX  
VITIS™

Executable  
可执行

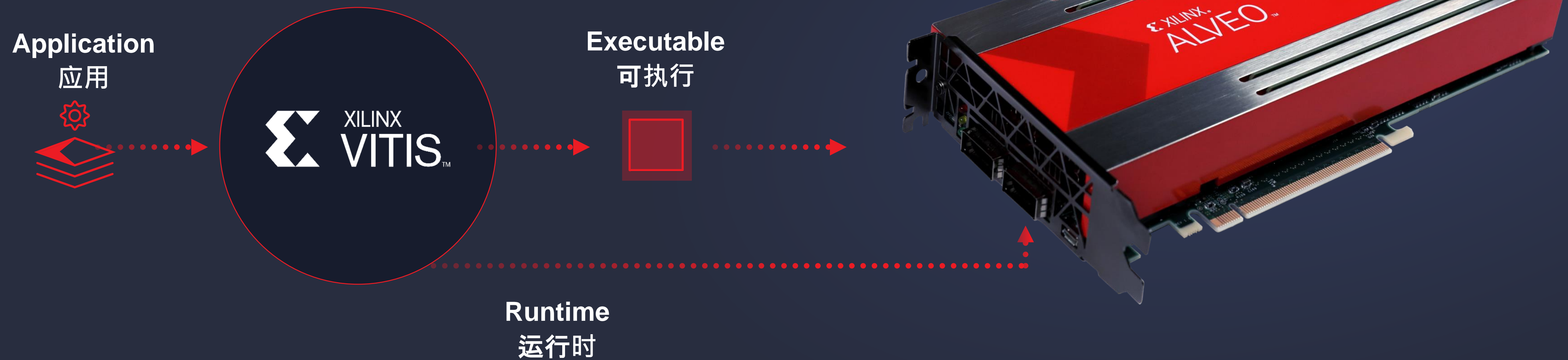


Runtime  
运行时



Deploy

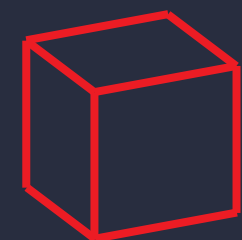
# Single Server Deployment 单服务器部署



Deploy

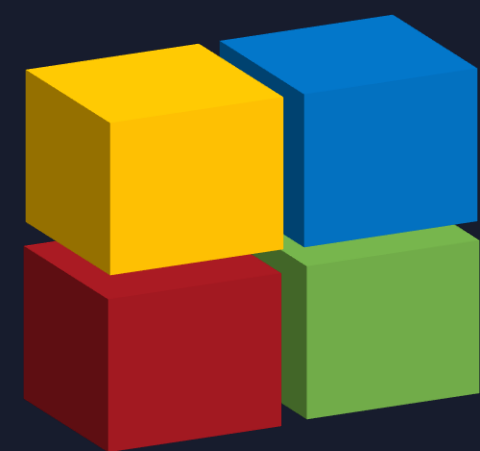
# Scale Out Deployment 扩展部署

Executable  
可执行

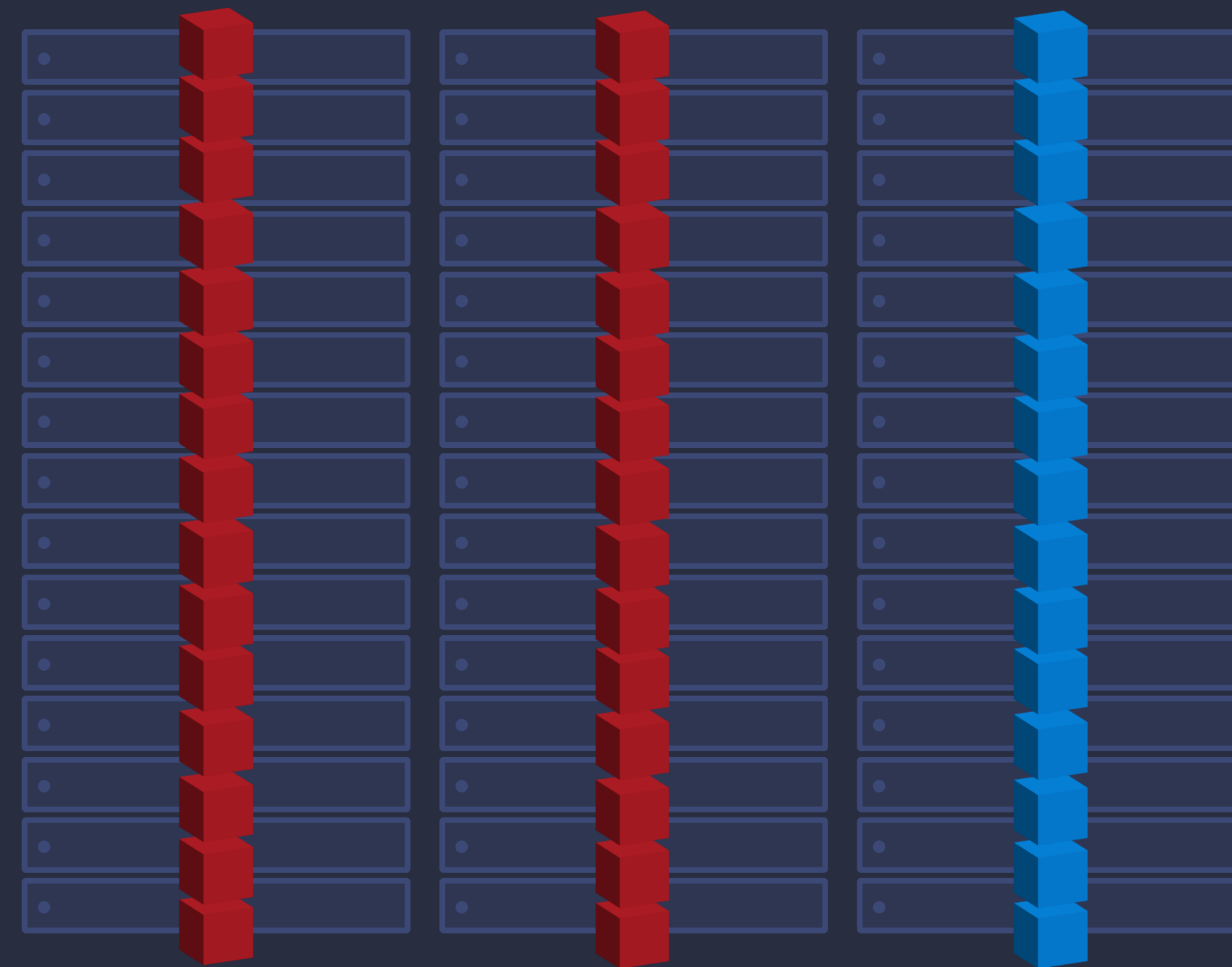
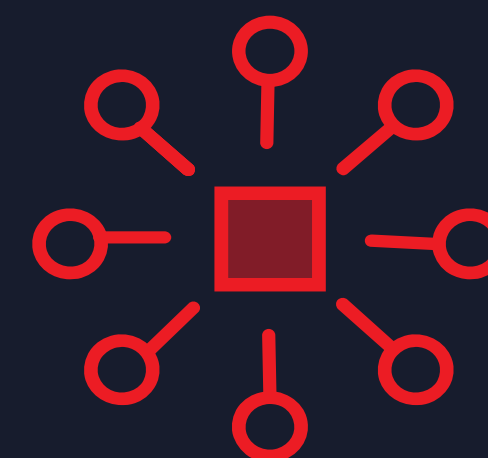


Runtime  
运行时

Xilinx Docker  
Registry

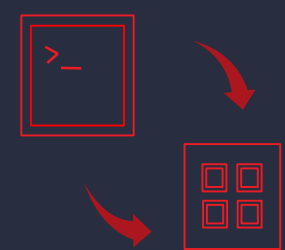


Scale Out



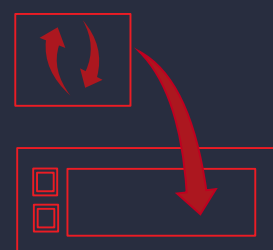
# Vitis AI: From TensorFlow to Implementation in Minutes

## Vitis AI : 几分钟实现 TensorFlow 部署



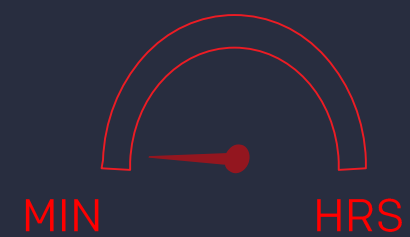
**DNN Processing Unit (DPU)**

DNN 处理器 (DPU)



**Direct Framework Compilation**

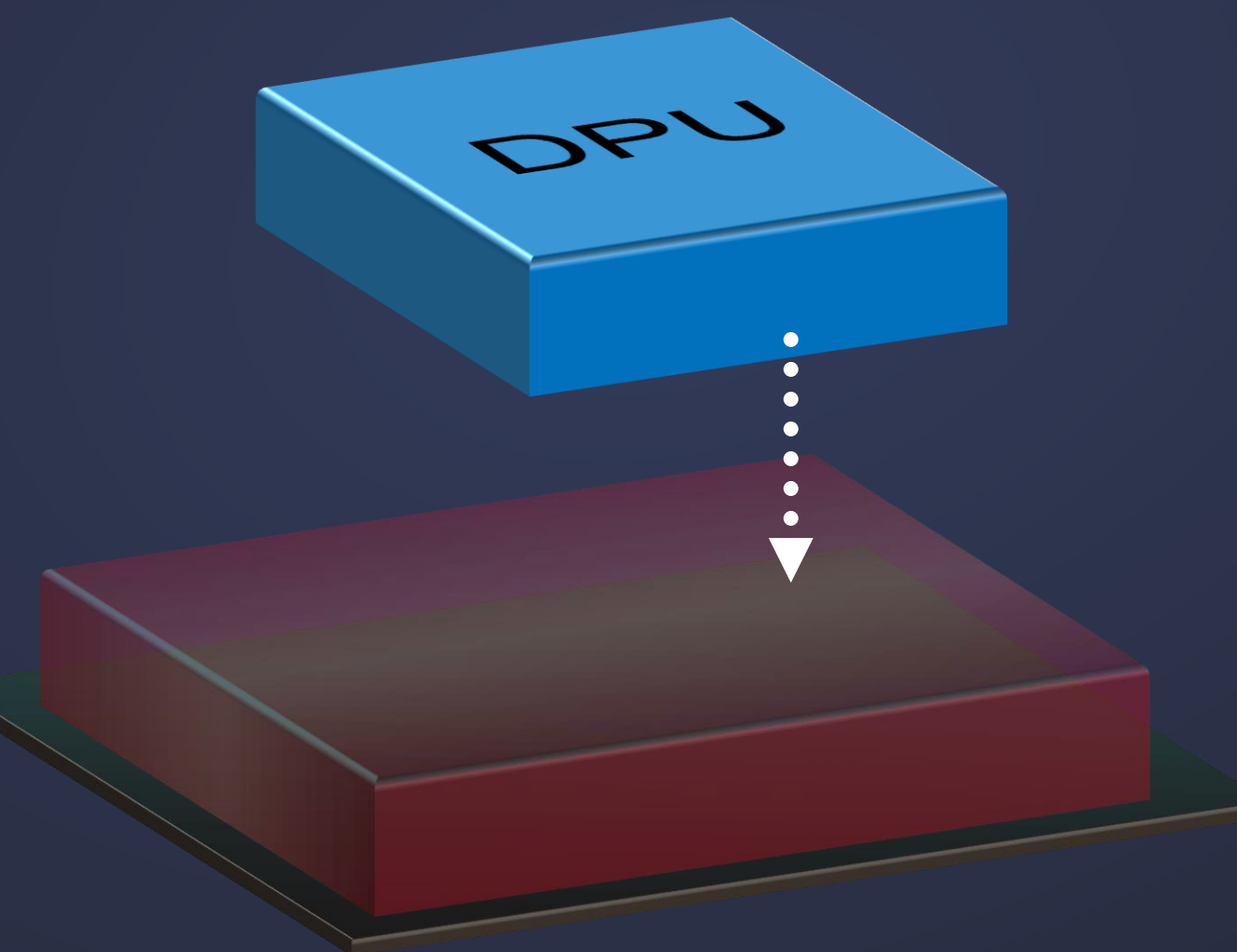
直接框架编译



**Minutes of Compile Times**

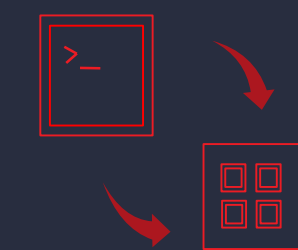
编译时间

**Adaptive Device**  
自适应器件



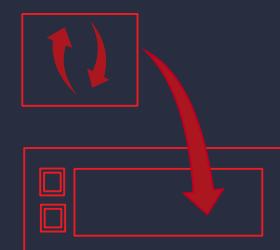
# Vitis AI: From TensorFlow to Implementation in Minutes

## Vitis AI : 几分钟实现 TensorFlow 部署



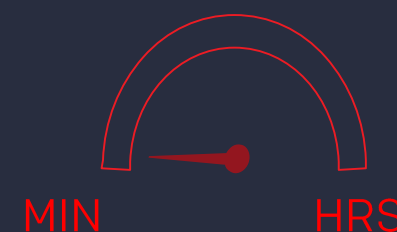
**DNN Processing Unit (DPU)**

DNN 处理器 (DPU)



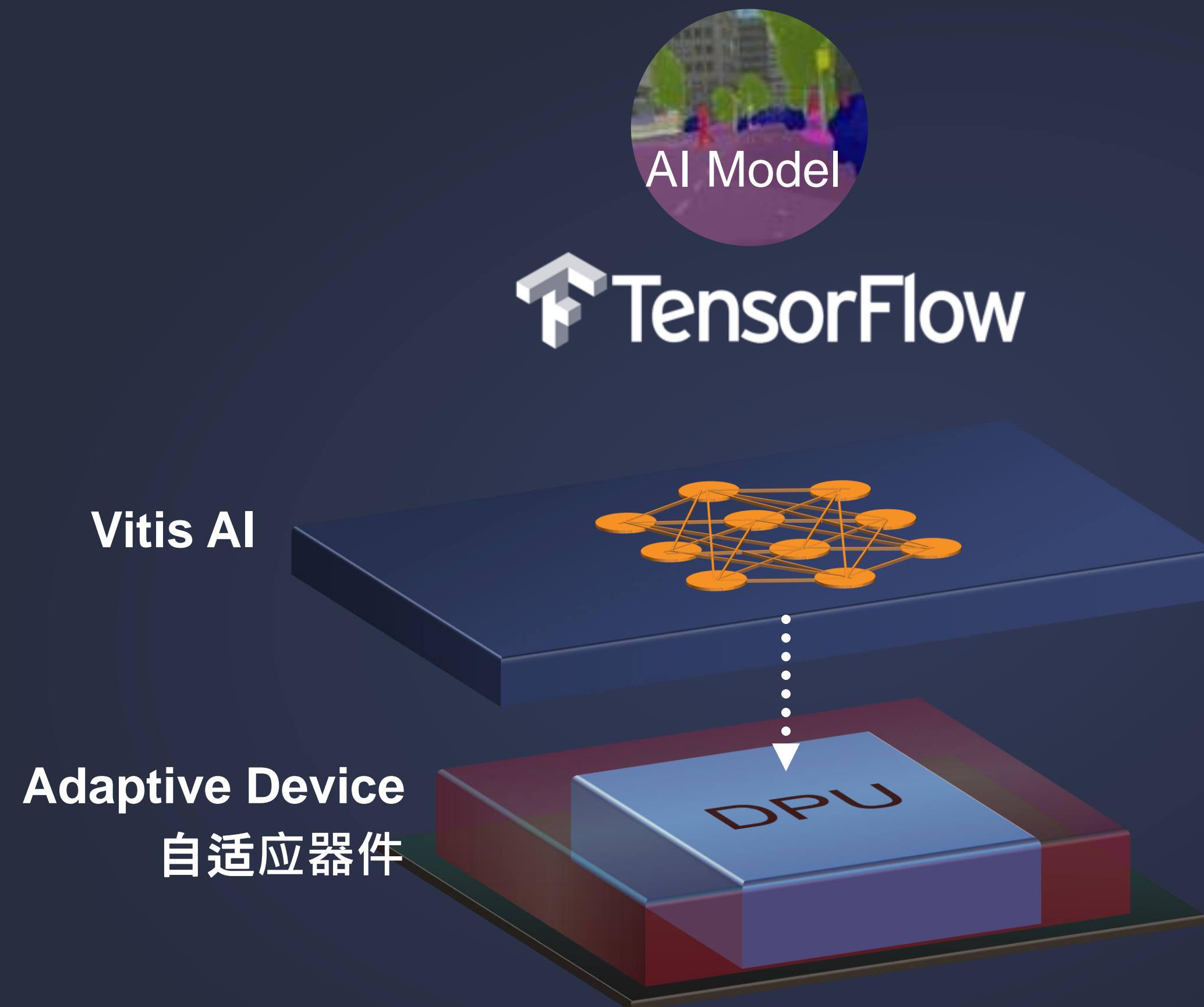
**Direct Framework Compilation**

直接框架编译



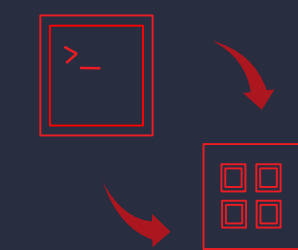
**Minutes of Compile Times**

编译时间



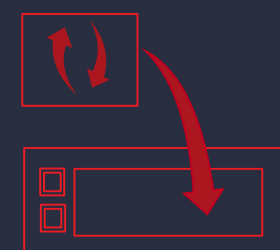
# Vitis AI: From TensorFlow to Implementation in Minutes

## Vitis AI : 几分钟实现 TensorFlow 部署



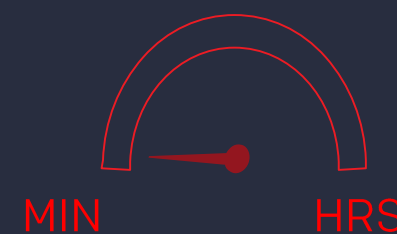
**DNN Processing Unit (DPU)**

DNN 处理器 (DPU)



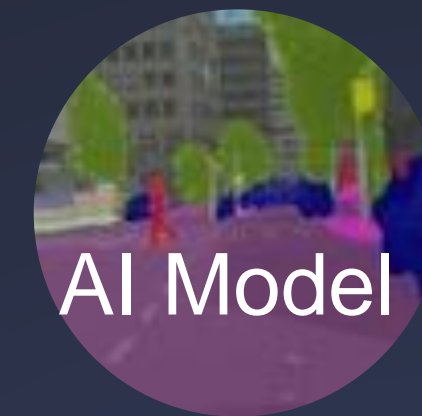
**Direct Framework Compilation**

直接框架编译



**Minutes of Compile Times**

编译时间



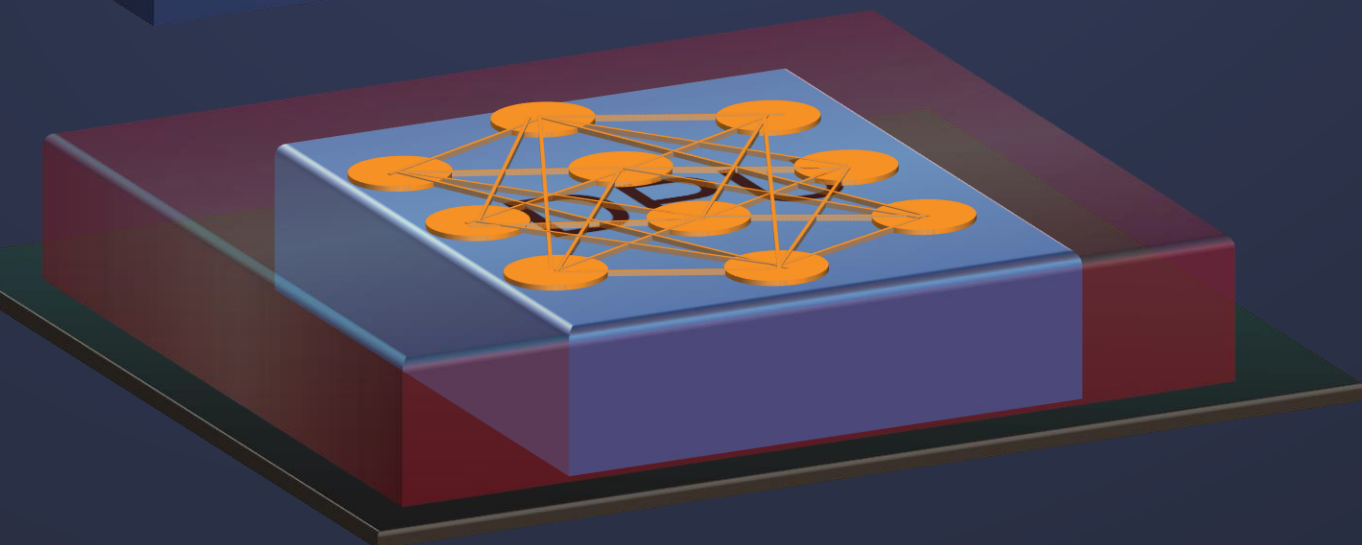
**TensorFlow**

**Vitis AI**



**Adaptive Device**

自适应器件



# Enabling AI 加速 AI 实现

Frameworks  
框架

TensorFlow

Caffe

PyTorch

Vitis AI  
Models  
Vitis AI 模型



Vitis AI  
Development Kit  
Vitis AI 开发套件

AI Optimizer  
AI 优化器

AI Quantizer  
AI 量化器

AI Compiler  
AI 编译器

AI Profiler  
AI 分析器

Vitis Drivers & Runtime (XRT)  
Vitis 驱动及运行时 (XRT)

DSA

CNN DPU  
卷积神经网络 DPU

LSTM DPU  
长短期记忆 DPU

MLP DPU  
机器学习处理器 DPU

60+ pretrained, optimized reference models  
60 多个预训练、优化的参考模型

Performance improvement up to 10-20x  
性能提升高达 10-20 倍

Tensor based ISA for true software programmability  
基于 Tensor 的指令集架构 (ISA) 实现真正的软件可编程

**Announcing:  
Vitis AI Now Available  
For Download**

**今日新闻：  
Vitis AI 现可  
免费下载了！**





# developer.xilinx.com

## 赛灵思开发者社区

Beta Site Now Available 网站公测版现已推出

Tutorials, Articles & Projects 教程、文章和项目

Accessible from a Single Location 一站式获得所有内容

Learn Directly from Vitis Experts 直接向 Vitis 专家学习

Model	FPS
FPN	64.29
Unet-Full	14.27
Unet-Lite	36.25

Using data gathered throughout this tutorial, we can compare the performance of the ZCU102 vs. the GTX1080ti graphics from section 4.1. Albeit, this isn't a fair comparison for two reasons:

1. We are comparing an embedded ~20W device with a 225W GPU
2. The ZCU102 execution time includes reading/preparing the input and displaying the output whereas the GPU measures inference time of the models

That said, this still provides some data points which are useful to garner further understanding. The following chart shows values measured on the ZCU102 vs. the GTX1080ti.

Model	ZCU102 (FPN)	GTX1080ti (Floating Point)
Enet	~60	~35
ESPNet	~40	~95
FPN	~65	~65
Unet-Full	~15	~30
Unet-Lite	~35	~70

What is perhaps a bit more useful than comparing raw FPS, however, is to compare FPS/W (performance/Watt) as this is performance is achievable for a certain power cost. Bear in mind, this is still not a fair comparison due to reason 2, but the little more in this light. In reality the advantage is even more pronounced if only the DPU throughput is considered.

In order to perform this comparison, ~20W was measured on the ZCU102 board during forward inference, and the nvidia-smi during forward inference of each of the models as part of section 4.1. The comparison between the two can be seen in the following chart.

Integrating optimized RTL Kernels into Accelerated Applications using Vitis

Ted Ennis

Sep 24, 2019 3:11:58 PM

YOLO: Input Image → Color Conv → Resize → Scale → To CNN

GoogleNet: Input Image → Resize → Mean Subtraction → To CNN

Accelerating ML Preprocessing with Vitis Vision

Alvin Clark

Sep 24, 2019 2:59:04 PM

Profiling and Accelerating C++ Applications and Algorithms

Christophe Charpentier

Sep 24, 2019 2:57:45 PM

Sequential execution (two runs)

Task-level parallelism and pipelining in HLS (fork-join and beyond)

Frédéric Rivoallon

Sep 23, 2019 3:57:27 PM

Cosine Similarity Using Xilinx Alveo

Alvin Clark, Kumar Deepak, Liang Ma

Sep 24, 2019 1:13:57 PM

$$\text{similarity}(A, B) = \frac{A \cdot B}{\|A\| \times \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n A_i^2} \times \sqrt{\sum_{i=1}^n B_i^2}}$$

Sequential execution (two runs)

Task-level parallelism and pipelining in HLS (fork-join and beyond)

Frédéric Rivoallon

Sep 23, 2019 3:57:27 PM

Get Moving with Alveo: Example 8 Pipelining Operations with OpenCV

Rob Armstrong

Sep 23, 2019 3:31:04 PM

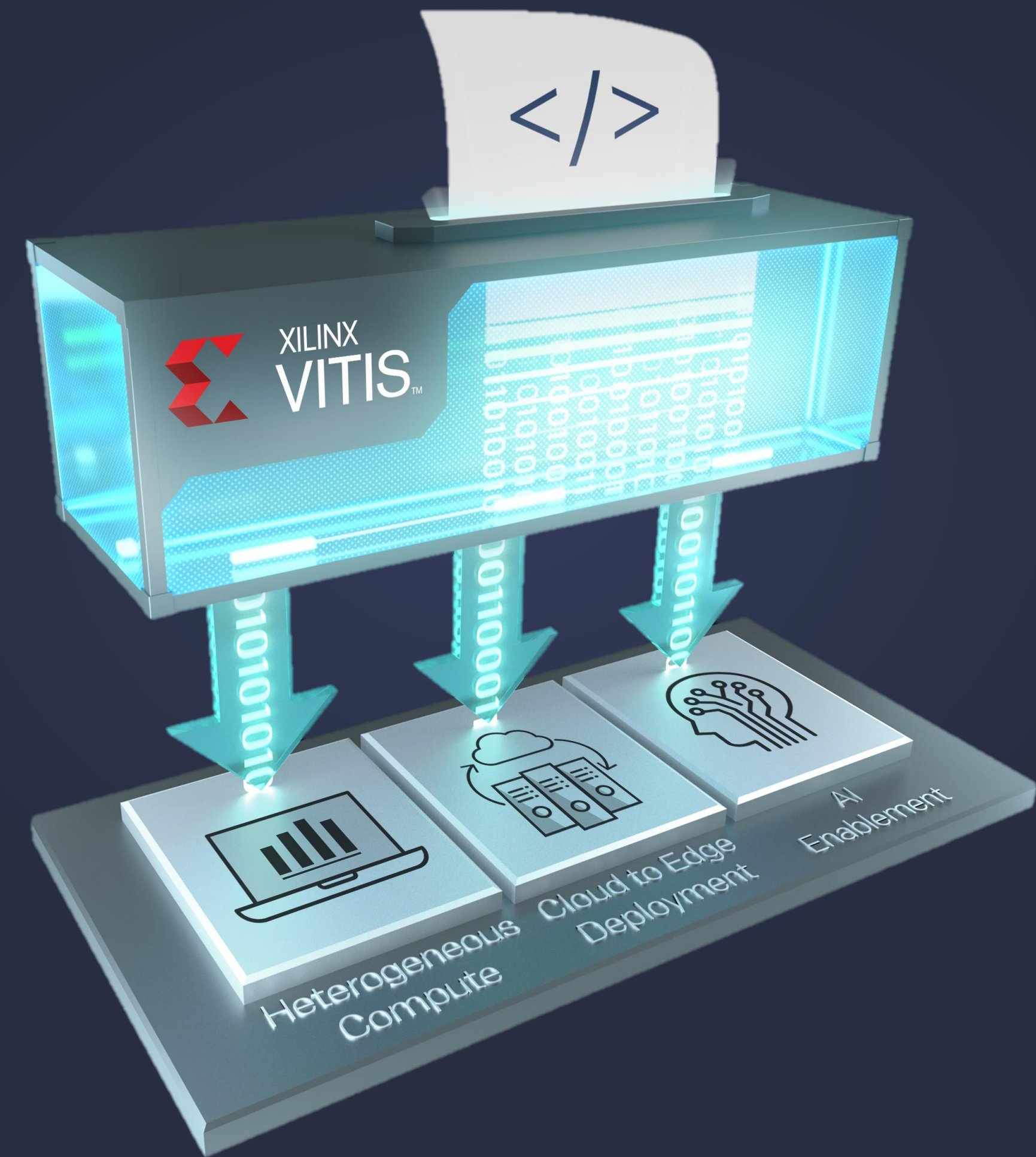
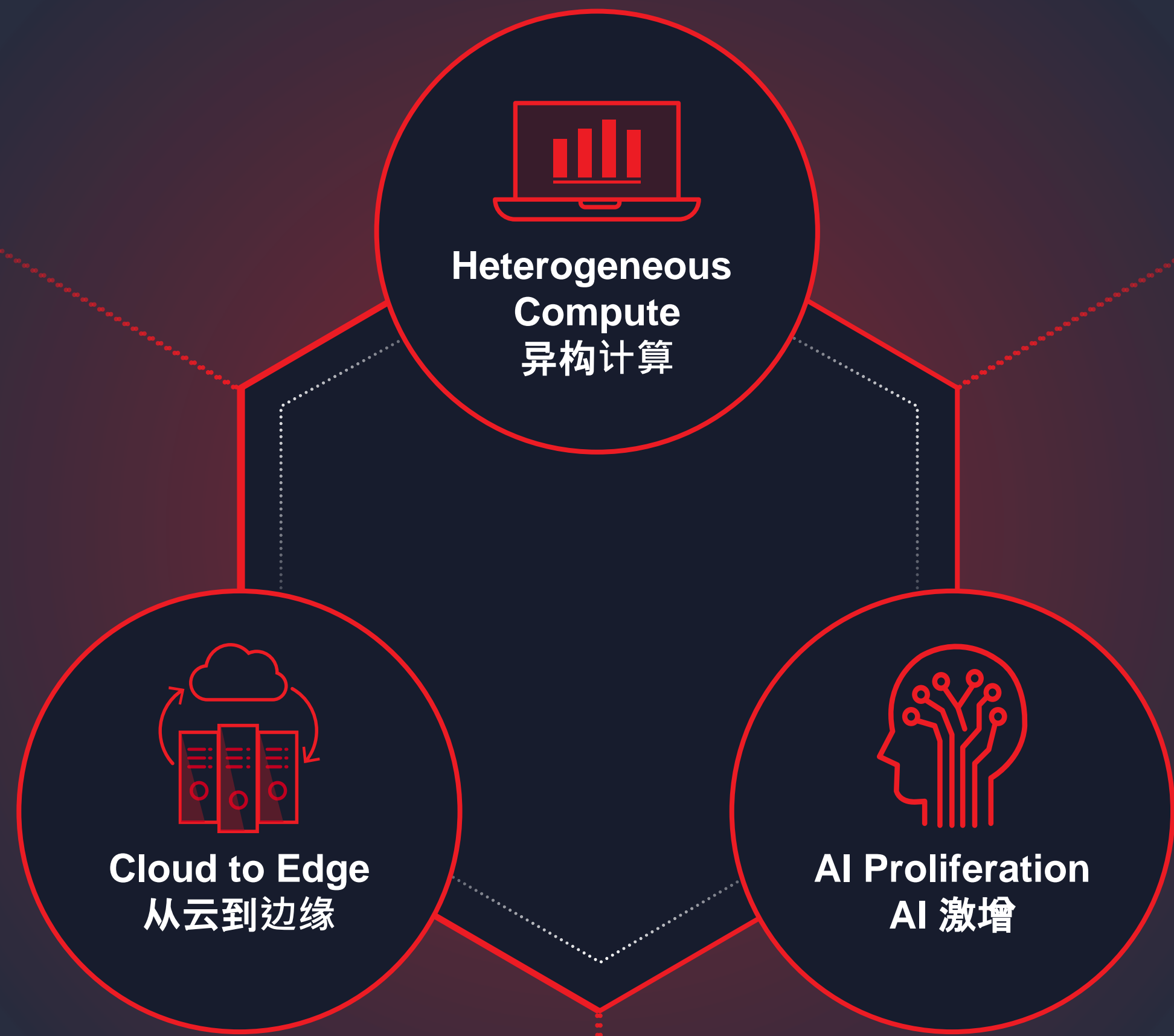
Get Moving with Alveo: Example 7 Image Resizing with OpenCV

Rob Armstrong

Sep 23, 2019 3:30:44 PM

### 50+ expert articles & projects (and growing)

### 50 多篇/个专家文章与项目 (且持续增加中)





Alibaba Cloud



阿里云



**Jeff Zhang 张振祥**

Alibaba Cloud and Intelligent FPGA  
Heterogeneous Computing Leader  
阿里云和智能 FPGA 异构计算负责人





Jackie Cheng 程龙

Zhongtai Securities, CIO

中泰证券， CIO



百度大脑EdgeBoard加速产业智能化升级

Hongguang Zhang 张红光

Technical Manager 技术经理

Baidu 百度

# Development Platforms for ALL Developers 适合所有开发者的开发平台

