解决方案简介

FPGA
实时分析流数据

Megh Computing 平台使用英特尔® Arria® 10 FPGA 加速实时流分析

使用机器学习和深度学习从流数据中无缝提取更多价值

“数据中心的第三个计算浪潮已经拉开序幕。现在，基于 FPGA 的专用硬件加速器支持在内部和云中使用，以满足分析和其他应用的实时流传输、机器学习和深度学习需求。”

— Megh Computing 首席执行官 Prabhat K. Gupta

要点概述

由于来自传感器、网络和其他来源的流数据的爆炸性增长，实时流处理的需求正在迅速增加。企业想通过处理流数据来创造业务价值。越来越多的企业采用包含机器学习和深度学习库的 Spark Streaming* 框架，从流数据中提取信息。如今，完全基于软件的解决方案无法满足低延迟、高吞吐率的性能要求。必须使用全新的集成硬件和软件的平台与工具，才能提供所需的低延迟和高计算能力。

Megh Computing 提供了一个使用 Spark Streaming 和其他框架加速实时分析的创新平台。该解决方案在基于英特尔® Arria® 10 FPGA 的硬件和软件平台、英特尔® 可编程加速卡（英特尔® PAC）和面向采用 FPGA 的英特尔至强® CPU 的英特尔加速堆栈上运行，可无缝地加速使用机器学习和深度学习算法处理数据流（以提取价值）的应用。借助 Megh 和英特尔解决方案，组织可以提升性能，降低延迟和总体拥有成本，同时支持新的用例和工作负载。

挑战

预计到 2022 年，流分析市场规模将达到 159 亿美元，复合年增长率为 33.1%。[1] 推动这一增长的主要因素是结构化和非结构化数据流的激增，这些数据主要来自物联网传感器和事件、社交媒体、追踪数据使用和行为的 web 应用以及多个垂直细分市场的交易和运营数据。企业需要从依赖传统商业情报（BI）转型为通过机器和深度学习获得高级分析。由于大量流数据需要更高的性能、更低的延迟和加速处理，这对计算和基础设施提出了更高的要求。

当今的实时分析解决方案基于软件平台，使用 Apache Kafka*、Apache Flink* 或类似的框架传输数据流，或将 Spark Streaming 框架用作处理数据的分布式框架。典型的运营阶段包括（1）提取、转换、加载（ETL），用于提取数据并为处理做好准备，（2）机器学习和/或深度学习库，用于从数据中推断信息。

这些开源或完全基于软件的专用解决方案无法满足实时分析日益增长的计算需求，也无法提供支持实时分析用例所需的低延迟。此外，为了处理激增的数据，组织扩大了节点数量，但是它们发现性能并没有线性扩展。

一般情况下，仅凭 CPU 无法在高速运行时降低延迟。FPGA 可以用作异构 CPU 和 FPGA 平台的一部分，来实施可配置的应用专用加速器并提供更高的性能和更低的延迟。FPGA 比 GPU 或其他加速器更适合此类应用，因为它们具有加速性能、高灵活性、低功耗、可编程性等优势。不足之处在于，高效编程和管理 FPGA 比较复杂，并且需要集成到数据中心编排软件。

当前的解决方案使用包含机器学习和深度学习库的 Spark Streaming*，通常无法满足性能要求

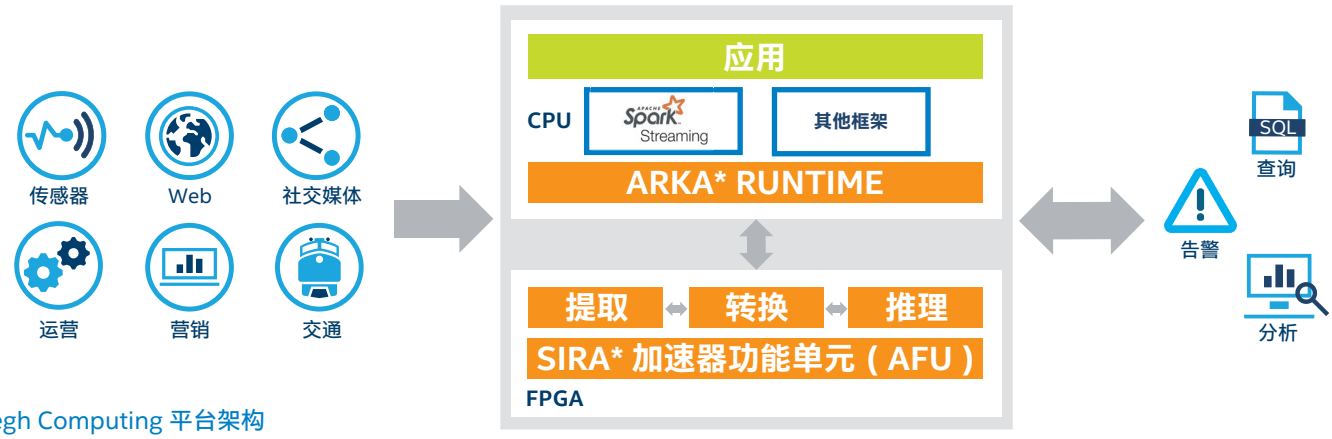
解决方案

Megh Computing 实时分析平台能够应对实时分析的关键挑战，支持以各种格式流传输数据，包括音频、视频和文本。

Megh 解决方案基于异构 CPU+FPGA 平台，采用搭载英特尔® Arria® 10 FPGA 可编程加速卡（PAC）和英特尔加速堆栈的标准大容量英特尔® 至强® 服务器。Megh 解决方案拥有 3 个主要属性：

- 在 FPGA 中实施完整的实时分析管道：
  提取： 用于以不同的格式流传输数据
  转换： 用于转换数据以进行处理
  推理： 用于机器学习和深度学习推理
- 支持包含机器学习和深度学习库的标准和定制数据分析框架，无需修改代码。例如，使用 Spark Streaming 和 BigDL* 库进行推理。
- 使用面向 Megh Arka Runtime 软件和 Sira* FPGA 硬件库的 SDK 定制不同用例。

在 Megh 和英特尔的协助下，您可以获得堪比 ASIC 的易用性，同时获得 FPGA 的可编程性和重配置优势。英特尔® 平台上的 Megh 集成软件和硬件解决方案通过高级 API 提供预优化库，降低了 FPGA 的复杂性，本质上处理库调用并将其重新映射至 FPGA。英特尔 Arria 10 FPGA 直接处理内嵌流数据，将 CPU 留给其他数据中心活动使用。



Megh Computing 平台架构

Megh Computing 实时分析平台优势	
提高性能和效率	• 总体运营成本降低 **50% 以上**[2] • 性能和效率提升 **5 至 10 倍以上**[2] • 延迟降至 **10% 以下**[2]
更快地创造价值	• 支持通过大容量硬件平台部署全新实时分析用例 • 获得开源软件框架的灵活性和定制 Megh SDK 的能力
简化部署	• 在公有云、私有云或边缘云中使用容器或虚拟机（VM）部署 FPGA 加速器即服务，并集成云编排层

面向数据经济的 FPGA

灵活的可编程 FPGA 使企业和行业能够满足不断增长的横向扩展加速器需求，使用机器和深度学习算法实时流传输分析。

FPGA 提供了大量的可配置硬件门，用于：

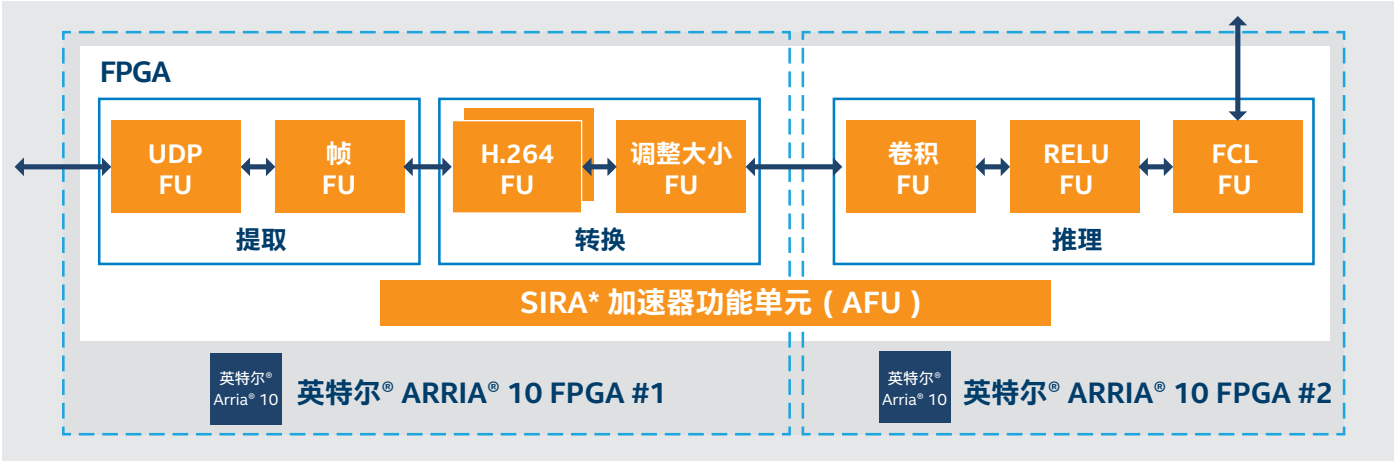
- 设计具有直接 I/O 连接和低延迟的定制硬件加速器
- 提高单个应用的性能和效率
- 快速将器件重新配置为不同应用的新加速器

Megh Computing 解决方案利用这些功能在 FPGA 中实施完整的实时分析管道。例如，Megh 创建了针对欺诈防范和其他零售市场用例的视频分析管道。完整的管道包括用于实施提取、转换和推理阶段的各种功能单元（FU）。这些功能映射至两个英特尔 Arria 10 FPGA：

- 提取和转换阶段在第一个 FPGA 中实施
- 推理阶段在第二个 FPGA 中实施

数据直接从第一个 FPGA 传输至第二个 FPGA。所有功能都可以被映射至单个英特尔® Stratix® 10 FPGA。解决方案也可以使用两个英特尔 Stratix 10 FPGA 来提供更大的容量。

近期，Megh Computing 将构建用于文本分析的类似管道，然后构建语音和其他管道。所有管道将利用 Megh 提供的加速功能单元（AFU）库和/或第三方开发的集成 FU。



映射至英特尔® Arria® 10 FPGA 的 Megh Computing 视频分析管道

示例用例

大量流数据和事务处理是金融、零售、电信、制造业等领域运营和决策的核心，也是现代企业的关键。无论在边缘还是云，Megh 和英特尔解决方案适用于需要低延迟、高性能的用例，能够快速传输、处理与分析流数据。

例如，金融交易分析必须满足 100 毫秒以内的实时要求。股市波动必须实时可见，并基于准确的最新风险评估调整投资组合。

欺诈防范对于金融和零售业至关重要，需要毫秒级的交易验证。电子商务和金融公司通过监控机器驱动的算法来发现异变，确保欺诈一出现便被检测到。在零售业，销售点（POS）终端必须实时检测不准确的 SKU 扫描或标记。

物联网正在边缘创建众多新用例，包括实时对象、图像和面部识别、增强和虚拟现实以及需要毫秒级响应时间的计算机视觉应用，以便为最终用户提供合格的体验。

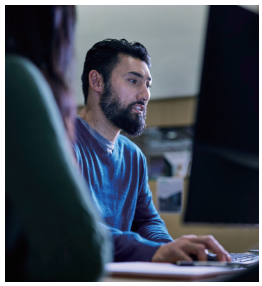
大量应用使用实时数据更新仪表盘，它们需要将响应时间控制数次秒以内，然而许多传统运营流程需要近乎即时的响应时间。

零售

- 防止错误扫描 SKU 导致的欺诈
- 对生鲜产品进行图像分类，以便控制质量
- 通过避免产品误放改进库存管理

金融

- 信用卡交易中的欺诈防范
- 实时监控交易

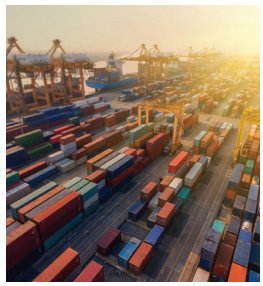


电信

- 基于网络探头分析与调整网络参数
- 实时监控网络数据，以进行安全检查

工业制造

- 实时监控生产线，以提高运营效率
- 预测性资产管理



企业

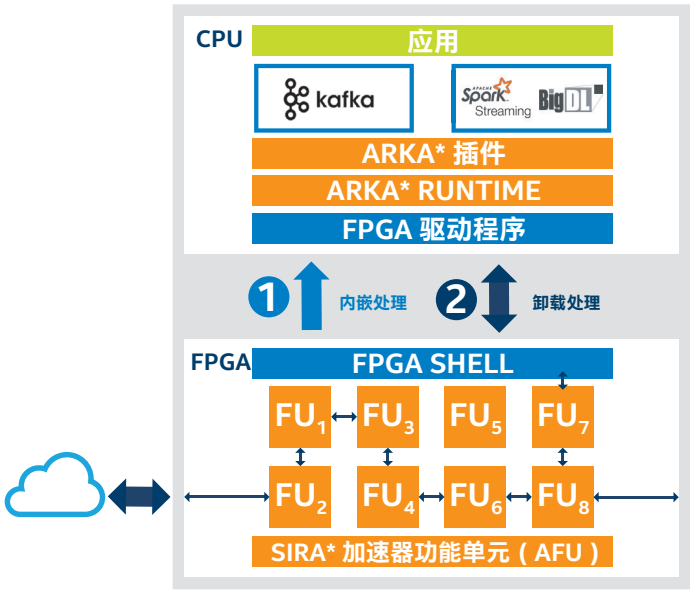
- 安全事件管理
- 分析实时和历史数据，以改进决策过程

解决方案架构

基本上，Megh Computing 解决方案通过使用开源和定制框架的插件库，降低 FPGA 的复杂性。

Megh 将软件和硬件相结合，提供了对流数据进行实时分析处理所需的重要性能和灵活性。硬件平台由采用英特尔 Arria 10 或英特尔 Stratix 10 FPGA 的英特尔® 可编程加速卡（英特尔® PAC）组成，在基于英特尔® CPU 的服务器上运行。对于软件，包含 BigDL 库的 Spark Streaming 框架与面向 Arka Runtime 和 Sira 加速器功能单元（AFU）的 Megh 库相结合。

- Arka Runtime 提供加速器即服务功能，管理英特尔® FPGA 并支持 AFU 的软件回退。它还为框架构建定制数据流加速器管道和插件。
- Sira AFU 提供真正的加速并且作为库运行，可下载至英特尔 FPGA。应用使用标准或定制 API 照常运行。



Megh 平台充当应用和英特尔® Arria® 10 FPGA 之间的中间层，简化并加速了流数据的实时分析处理

1. 通过 Arka* Runtime 框架的高级 API 将高度优化的 Sira* 加速器功能单元提供给应用。
2. 应用可以照常运行并调用 API。

3. 将功能加载至英特尔® Arria® 10 FPGA 并在此处执行。

Megh 解决方案提供以下优势：

- 在 FPGA 中实施的完整实时分析管道
  - 固定的低延迟，可实施实时用例
  - 高吞吐率，有助于降低总体拥有成本
- 面向 FPGA 的软件抽象层，以简化使用不同框架的部署
  - 管理与配置 FPGA
  - 提供 FPGA 加速器即服务
- 协助快速上市的定制工具
  - 用于开发定制加速器和管道的 Arka 和 Sira SDK
  - Megh Computing 提供的支持服务

英特尔 Arria 10 和英特尔 Stratix 10 FPGA

随着市场对大数据和人工智能需求的增长，FPGA 的可编程技术应满足数据中心应用的处理要求，适应其不断变化的工作负载。借助可再配置逻辑、内存和数字信号处理模块，FPGA 可通过适当编程执行任何类型的功能，同时实现高吞吐量和实时性能，因而适用于许多关键的企业和云应用。

面向包含英特尔 FPGA 的英特尔® 至强® 处理器的加速堆栈支持行业领先的操作系统、虚拟化和编排软件，能够为软件开发人员提供一种通用接口，帮助他们更快速获取收入、简化管理并利用日益扩大的加速工作负载生态系统。

与采用英特尔 Arria 10 FPGA 的英特尔 PAC 一样，采用英特尔 Stratix 10 FPGA 的最新英特尔 PAC 支持 IP 加速广泛的应用工作负载。采用 Stratix® 10 SX FPGA 的英特尔® PAC 是一种外形更大的卡，专为内嵌处理和内存密集型工作负载而构建，如流分析和视频转码。采用 Arria 10 FPGA 的英特尔 PAC 外形小巧，适用于回溯测试、数据库加速和图像处理工作负载。

展望未来

Megh Computing 实时分析平台针对通过英特尔 FPGA 实现低延迟和高吞吐率的用例。该解决方案专门面向标准化硬件和软件而设计，支持轻松地从一个节点扩展到大型集群，以及部署于公有云、私有云和边缘云。

在 Megh 和英特尔的帮助下，组织可以利用当前的基础设施，更快速地获得机器学习和深度学习的优势。

了解更多信息

如欲了解 Megh Computing 实时分析平台，请访问 megh.com 或拨打电话 +1（888）428.2396。

了解有关英特尔可编程加速卡和加速堆栈的更多信息。

解决方案组件

Megh Computing 实时分析平台和库

Arka* Runtime

- 提供与管理加速器即服务
- 支持连续数据流
- 支持多个 FPGA

Sira* 加速器功能单元（AFU）

- 数据包处理引擎（PPE）：UDP 数据包处理和 TCP 旁路
- 流处理引擎（SPE）：数据转换
- 深度学习引擎（DLE）：面向不同拓扑的内嵌推理

采用英特尔® Arria® 10 FPGA 的英特尔® 可编程加速卡（英特尔® PAC）

基于英特尔® 处理器的合格服务器

有关 Megh Computing

Megh Computing 的目标是使用基于 FPGA 的横向扩展加速器推动数据中心的第三个浪潮。2017 年，英特尔的一个团队创立了 Megh Computing，我们率先将 FPGA 应用于数据中心。在过去的十年间，团队承诺将使用 CPU 和 FPGA 平台进行异构计算的概念变为现实。我们现阶段的重点是利用我们在全堆栈开发方面的专业知识，交付一个在公有云、私有云和边缘云使用 FPGA 加速器加速实时分析的平台。

megh.com

MEGH COMPUTING

1. marketresearchandstatistics.com/ad/global-streaming-analytics-market

2. 仅限 CPU 的配置基于英特尔® 至强™ 处理器 8000 系统，后者包含内存和一个执行程序；46 个内核。使用包含 BigDL 库的 ResNet-50 模型以 182 帧/秒的性能进行推理的功耗预计为 400 W。FPGA 系统包括采用两块英特尔® Arria® 10 FPGA 卡的英特尔® 至强™ 处理器 3000 系统。在 FPGA 上运行的 ResNet-50 模型性能为 939 帧/秒时，CPU 的功耗预计为 150 美元/W，FPGA 的功耗预计为 3x 50W。

性能结果基于 2018 年 12 月的测试，可能无法反映所有公开可用的安全更新。请参阅配置披露了解详细信息。没有任何产品能保证绝对安全。

在性能测试过程中使用的软件及工作负载可能仅针对英特尔微处理器进行了性能优化。SYSmark 和 MobileMark 等性能测试使用特定的计算机系统、组件、软件、操作和功能进行测量。上述任何要素的变动都有可能导致测试结果的变化。您应当参考其它信息和性能测试以帮助您完整评估您的采购决策，包括该产品与其它产品一同使用时的性能。更多信息请访问：https://www.intel.cn/content/www/cn/zh/benchmarks/benchmark.html

英特尔技术的特性和优势取决于系统配置，可能需要支持的硬件、软件或服务激活。实际性能可能因系统配置的不同而有所差异。任何计算机系统都无法提供绝对的安全性。请联系您的系统制造商或零售商，或访问：http://www.intel.cn/content/www/cn/zh/homepage.html

0219/MG/CMD/PDF 338557-001CN