

## IPU-POD<sub>4</sub> DA

为致力于实现AI新突破的创新者设计



### 开始探索新的AI边界

IPU-POD4 DA (Direct Attach, 直连) 是进行探索、创新和开发的理想平台。它能够帮助AI团队在机器智能方面取得新的突破。由主机服务器支持的1个IPU-M2000在价格合理的紧凑型2U系统中提供强大的1 petaFLOPS AI计算，可用于训练和推理工作负载。

### 直接连接即插即用

IPU-POD4 DA旨在让您立即启动并运行。作为一个开箱即用的系统，IPU-POD4 DA是一个完整且适配完成的AI计算系统，已为在数据中心安装准备就绪。Graphcore的AI专家和精英合作伙伴网络可为您提供广泛的文档和支持。

### 从小开始，扩展至大

IPU-POD4 DA是强大的独立AI计算资源。但它也为您提供了根据自身条件不断发展的机会。您投资的IPU-POD4 DA系统可以在日后扩展为更大的IPU-POD系统。

### 为扩展而建的AI基础架构

IPU-Fabric专门针对大规模AI工作负载的通信需求而设计，它采用了行业标准的IT设备，是Graphcore的创新型、低时延、无抖动互连。无论系统大小如何，它可在整个系统中支持高效、确定性的全局IPU互连。

## 系统技术参数

IPUs	4个GC200 Mk2 IPU
IPU-M2000s	1个IPU-M2000
Exchange-Memory	131.6GB (包括3.6GB处理器内存存储和128GB流存储)
性能	1 petaFLOPS FP16.16 250 teraFLOPS FP32
IPU核数	5,888
线程数	35,328
IPU-Fabric	2.8Tbps
Host-Link	100 GE RoCEv2
软件	Poplar
系统重量	16.5公斤+主机服务器
系统尺寸	2U
主机服务器	Graphcore合作伙伴提供多种经认证的主机服务器选项
散热	空气冷却
可选交换版本	请联络Graphcore销售人员

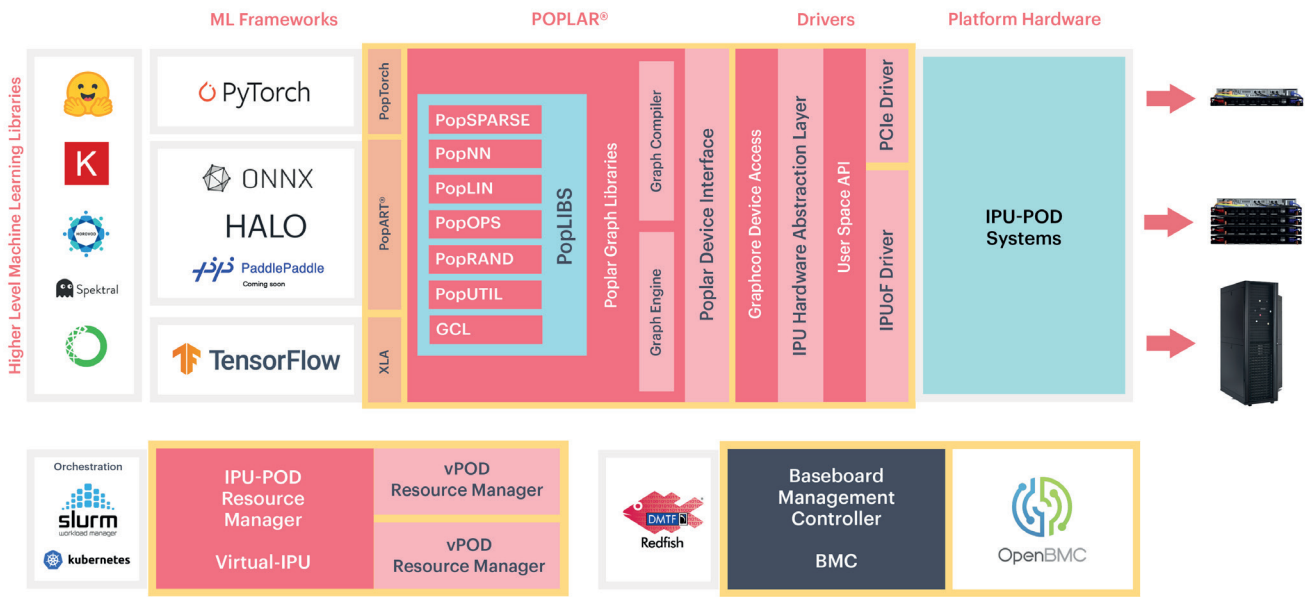
# GRAPHCORE 拟末

## 软件为先

我们的Poplar软件完全集成并且经过IPU优化，利用IPU架构的独特特性来构建具有无与伦比的性能和灵活性的AI应用程序。在不增加开发复杂性的前提下，Poplar可轻松地将模型从一个IPU扩展到数百个IPU，从而使您可以专注于应用程序的准确性和性能。

## 为AI开发人员而创建

支持TensorFlow、PyTorch和其他被广泛使用的机器学习框架，并且开源，可与完整的PopLibs库一起用于社区驱动的协作和创新。针对希望完全控制以挖掘最大性能的开发人员，Poplar启用了C++中的直接IPU编程。



## 为易于部署而创建

带有Poplar SDK工具和 框架图像的预构建Docker容器可以让您快速启动并运行。IPU-POD4 DA具有易于使用且直观的网页GUI (图形用户界面)，可简化IPU资源的管理。在这里，您可以管理状态、执行系统测试，以及为工作负载配置IPU。

## 获取AI专业知识

Graphcore AI专家和我们的精英合作伙伴网络在全球范围内为安装、生产部署和应用程序开发提供丰富的经验和支 持。

## 准备好新一代体验机器智能吗？

联系以下合作伙伴，评估您的AI基础设施要求和解决方案适用性。还有问题需要解答？请直接通过[info\\_china@graphcore.ai](mailto:info_china@graphcore.ai)与Graphcore取得联系。

## GRAPHCORE.CN