GRAPHCORE

# IPU-POD$_{16}$ DA

## For Innovators making new AI Breakthroughs

### Start exploring new AI horizons

IPU-POD$_{16}$ DA (Direct Attach) is the ideal platform for exploration, innovation and development. This lets AI teams make new breakthroughs in machine intelligence. Four IPU-M2000s, supported by a host server, deliver a powerful 4 petaFlops of AI compute for both training and inference workloads in an affordable, compact 5U system.

### Plug-and Play with Direct Attach

IPU-POD$_{16}$ DA is designed to get you up and running in no time. A turnkey system, featuring IPU-M2000s directly attached to an approved host server ready for installation in your datacenter. Extensive documentation and support is provided both by AI experts at Graphcore and our elite partner network.

### Start small, scale big

IPU-POD$_{16}$ DA is a powerful, standalone AI compute resource. However, it also offers the opportunity for growth, on your terms. Your IPU-POD$_{16}$ DA system investment can be expanded later into a larger IPU-POD system.

### AI infrastructure built to scale

Designed specifically for the communication requirements of AI workloads at scale, IPU-Fabric is Graphcore's innovative low-latency, jitter-free interconnect using industry standard IT equipment. It supports highly efficient, deterministic, all-to-all IPU interconnect across your system regardless of size.

## System Specifications

| | |
|---|---|
| **IPUs** | 16 x GC200 Mk2 IPUs |
| **IPU-M2000s** | 4 x IPU-M2000 |
| **Exchange-Memory** | 526.4GB (includes 14.4GB In-Processor Memory and 512GB Streaming Memory) |
| **Performance** | 4 petaFLOPS FP16.16<br>1 petaFLOPS FP32 |
| **IPU Cores** | 23,552 |
| **Threads** | 141,312 |
| **IPU-Fabric** | 2.8Tbps |
| **Host-Link** | 100 GE RoCEv2 |
| **Software** | Poplar |
| **System Weight** | 66 kg + Host server |
| **System Dimensions** | 5U |
| **Host server** | Selection of approved host servers from Graphcore partners. |
| **Thermal** | Air-Cooled |
| **Optional Switched Version** | Contact Graphcore sales |

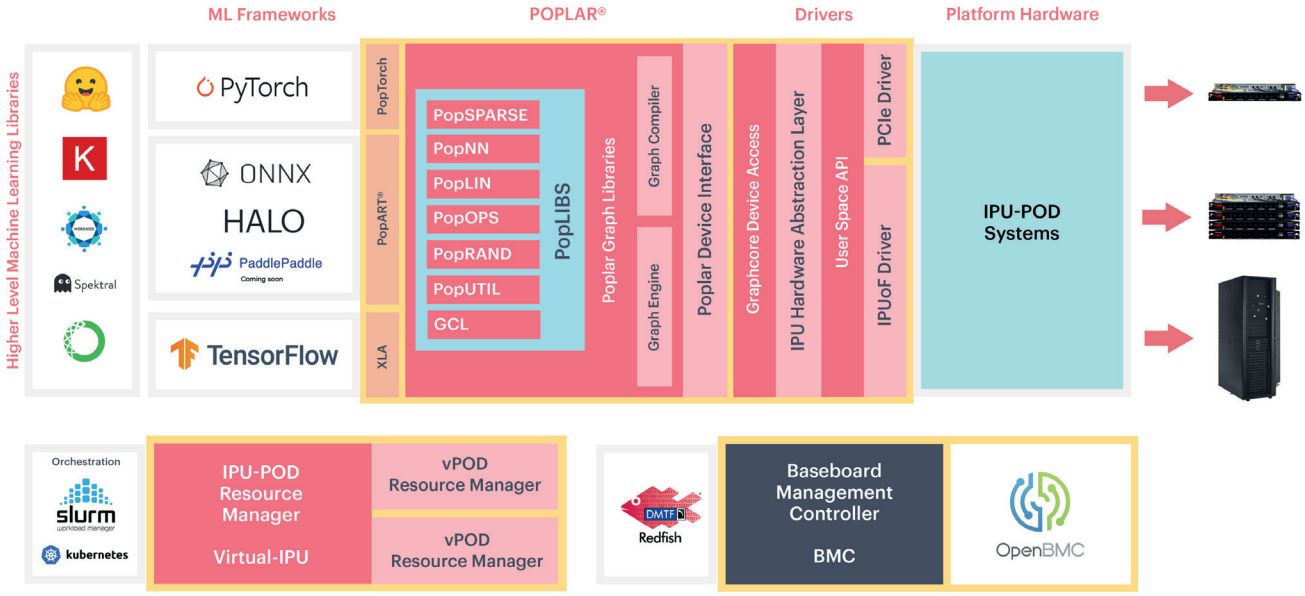
GRAPHCORE

## Software First

Fully integrated and IPU-optimised, our Poplar software leverages the unique characteristics of the IPU architecture to build AI applications of unrivalled performance and flexibility. Poplar allows effortless scaling of models from one to hundreds of IPUs without adding development complexity, allowing you to focus on the accuracy and performance of your application.

## Built for AI developers

TensorFlow, PyTorch and other popular ML frameworks are supported and available as open source, along with the comprehensive PopLibs library, for community driven collaboration and innovation. For developers who want full control to exploit maximum performance, Poplar enables direct IPU programming in C++.




## Built for deployment

Pre-built Docker containers with Poplar SDK tools and frameworks images let you get up and running fast. IPU-POD$_{16}$ DA has an easy-to-use, intuitive web GUI for simplified management of IPU resources. From here you can manage status, perform system tests, and provision IPUs for workloads.

## Access to AI expertise

A wealth of experience and support for installation, production deployment and application development is available globally from Graphcore AI experts and from our elite partner network.

## Ready to experience the next level in Machine Intelligence?

Connect with our partners below to assess your AI infrastructure requirements and solution fit.
Still have questions? Contact Graphcore directly at info@graphcore.ai

**GRAPHCORE.AI**