# 下一代gpu上Tensorflow性能的基准测试

随着机器学习（ML）研究人员和实践者不断探索深度学习的边界，对强大的GPU来训练和运行这些模型的需求越来越强烈。新的目标检测、图像分割和语音转录模型继续得到改进，在从自动驾驶到家庭助理的各种行业中都有应用。

为了满足这种对GPU计算的需求，亚马逊和谷歌最近都在其实例类型中添加了下一代Nvidia和GPU。另一家云GPU供应商，也提供了s。这些P100和Volta gpu是目前市场上最好的gpu，在ML应用程序的性能方面处于领先地位。这些gpu不仅比旧的gpu有更高的性能，而且还配备了16GB的内存，使ML模型更具表现力，并具有更大的训练小批量。



为了测试这些现代gpu如何执行典型的ML任务，我在Nvidia最新的gpu上训练了一个更快的R-CNN/resnet101目标检测模型。目标检测模型在300x300px图像输入上运行，训练的小批量大小为10、15和20个图像。

基准GPU：

* （16GB - 2.30美元/小时）
* （16GB - 1.73美元/小时）
* （16GB-$3.06/小时）
* （11GB - 个人电脑）

注：此\*\*\*\*基准测试专门针对较新的GPU，因此不包括较旧的K80和Quadro GPU。这些GPU是。

#### 结果

从性能的角度来看，Voltas无疑是当今最强大的gpu，尽管1080Ti只有9个月的历史，但它的性能远远超过Nvidia 1080Ti（约1.1-1.3x）和P100（约1.2-1.5x）。这延续了Nvidia发布越来越强大的GPU架构的快速节奏。

值得注意的是，亚马逊的Volta实例在同一个训练任务上的表现不如Paperspace的Volta。我自己对此的简短调查表明，可能是实例和GPU之间的I/O速度慢造成的；比较Amazon和Paperspace之间的纯GPU基准显示了类似的性能。

从成本的角度来看，PasSaleValas提供了很好的性价比，谷歌的P100的成本大约是10%，而亚马逊的售价则高达40%。

#### 我该用什么？

* 当然，大用户应该购买自己的gpu。从云提供商那里租用GPU仍然是一个昂贵的提议，而购买自己的GPU可以让你以更低的成本访问最好的硬件——前提是你可以利用它们来分摊成本。
* Paperspace Volta对于那些不想投入拥有自己的GPU的用户来说是一个很好的价值。对于只需要一个GPU的用户，使用Volta将提供良好的性能提升。
* 谷歌的P100S是最灵活的，允许用户将1, 2和4 P100GPU（或最多8 k80GPU）连接到任何实例，允许用户定制他们的CPU和GPU配置，以适应他们的计算需求。尽管P100的性能较差，但它们在成本调整的基础上也具有竞争力。
* 亚马逊的Voltas比谷歌的P100更强大，还提供附加1、4或8个gpu的能力。但是，用户无法自定义基本实例类型。在成本调整的基础上，它们也相对昂贵。我建议只有当你迫切需要8个GPU或者需要在EC2上时才使用它们。

Building something interesting? [Initialized Capital](http://twitter.com/@initializedcap) would love to chat with you.

**¹ Disclosure:**Paperspace is an Initialized Capital portfolio company.