

## Triton Users and Use cases Open Survey

This survey is intended to be filled by anyone that develops or uses Triton GPU kernels and related tools/infrastructure. The term “users” is very broad here, including everyone from developers to consumers of this technology. The results are public and the doc will be retained as public. It is hoped that the results will be valuable to everyone so participation in this survey is encouraged and appreciated.

Questions ? contact: [srampal@redhat.com](mailto:srampal@redhat.com)

<u>Your project/ organization</u>	<u>Where are you using Triton?</u>	<u>What problems is Triton solving for you? Any other feedback?</u>
Example: Acme corp  (Note: this row is just a dummy example to give you some ideas about the suggested kind of content you could add for each of your entries)	<ul style="list-style-type: none"><li>• We use Triton for rapid prototyping of new kernels for our custom multi-modal models across NVidia GPUs, AMD GPUs, our custom TPU-like accelerators. This is our x% use case</li><li>• We also use Triton kernels to selectively augment our inference service which is based on foo. This is our y% use case</li></ul>	<ul style="list-style-type: none"><li>• We do have limited CUDA programming expertise in-house in our org but it is not enough to keep up with our needs for quick kernel prototypes and Triton fits the bill perfectly to allow a larger set of developers build kernels for us</li><li>• In our experience Triton programmability x/10, performance y/10, open source community governance z/10</li></ul>
OpenAI		
AMD		
Meta		
Pytorch project - team X		
vLLM project - team Y		
LinkedIn		
Microsoft		
<add your row here>		