



Best practices, tutorials and ideas from Google Developer Experts (GDEs)

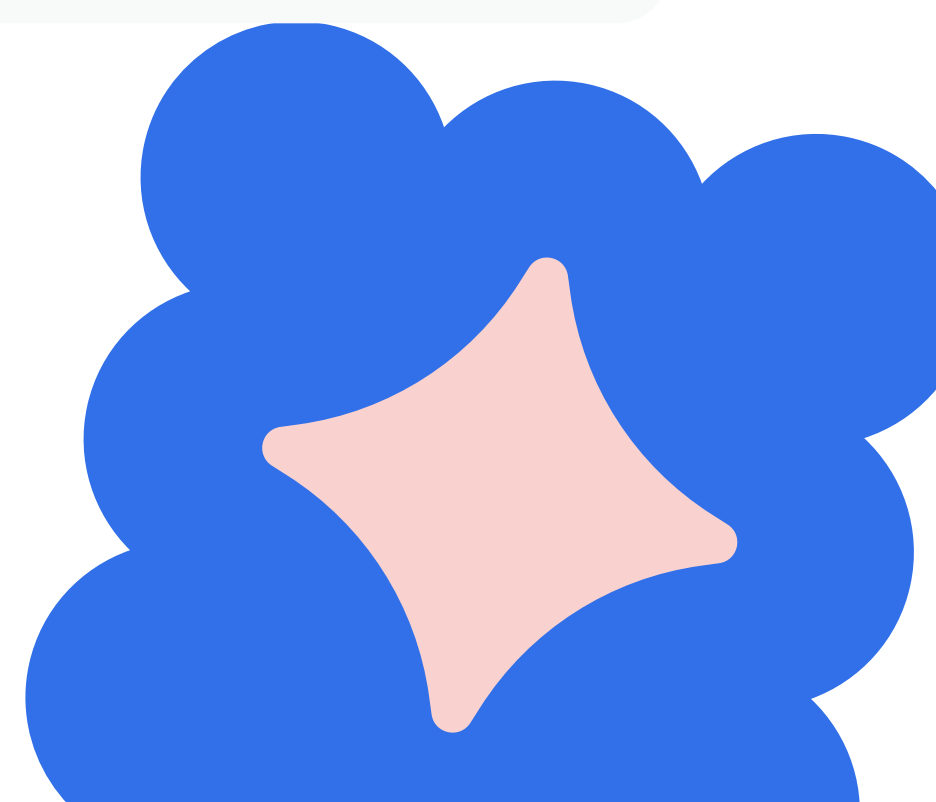
GDEs are a community of 1,000+ experts (customers, not Googlers) who have achieved the highest level of certification available. To become a GDE they must be nominated by a Googler and demonstrate deep technical expertise in Google Cloud. We've collated some relevant GDE content below:

Guide: JAX and Keras for LLM Development using TPU for Distributed Fine-Tuning:

In this tutorial, we'll dive into fine-tuning the Gemma 2B LLM using Keras, leveraging the power of JAX for high-performance computation, and distributing our training across a TPU device mesh for unparalleled speed and efficiency.

Implementing Neural Radiance Fields (NeRF) with Keras 3:

Historically, novel view synthesis was performed through hole-filled meshes or patchwork image blending. NeRF skips those processes by learning light interaction, geometry, and view-dependent effects in a single neural network. This produces photo-realistic novel views from sparse input—no stitch, no mesh, just neural rendering magic.



Learning resources

Documentation: Set up a Cloud TPU environment

Learn how to set up your Google Cloud project and your environment so that you can use Cloud TPU resources to train or run inference on models.

Documentation: Github recipes

Unlock recipes for reproducing workload performance on TPUs so that you can achieve your performance metrics.

Documentation: Quickstart guides

Don't want to start from scratch?
Choose from a repository of guides to help you start off on the right foot.

Documentation: How to scale your models

Demystify the science of understanding, scaling, and optimizing the performance of your models.

Conclusion

You now have a comprehensive understanding of TPUs. You learned:

- 1 How to effectively leverage Cloud TPUs for accelerating inference, including selecting appropriate options and maximizing their potential
- 2 About the diverse range of TPU accelerators available on Google Cloud, from foundational models like v2 and v3 to the latest innovations such as Ironwood (v7e) and Trillium (v6e), and the various scaling configurations, including TPU pods, slices, and multislice
- 3 Practical skills to deploy models and use powerful tools like the XProf profiler to monitor, diagnose, and tune their performance.

