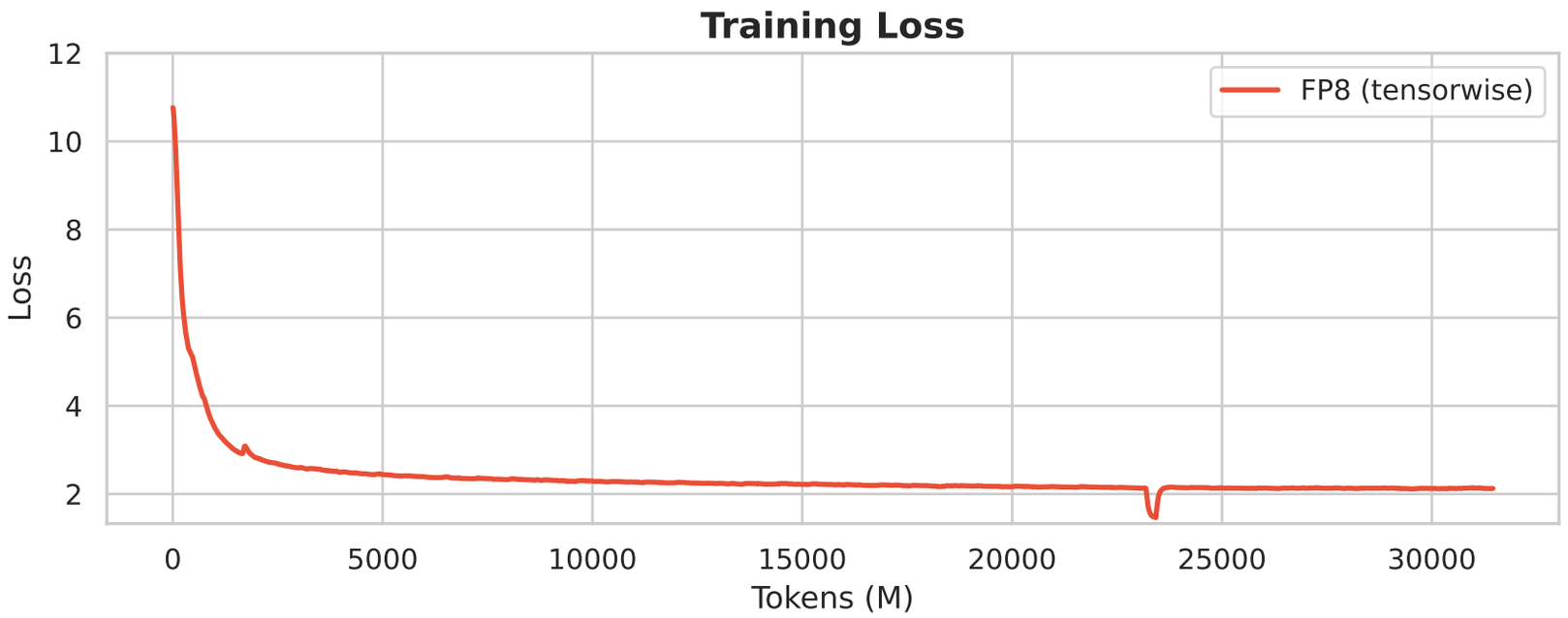
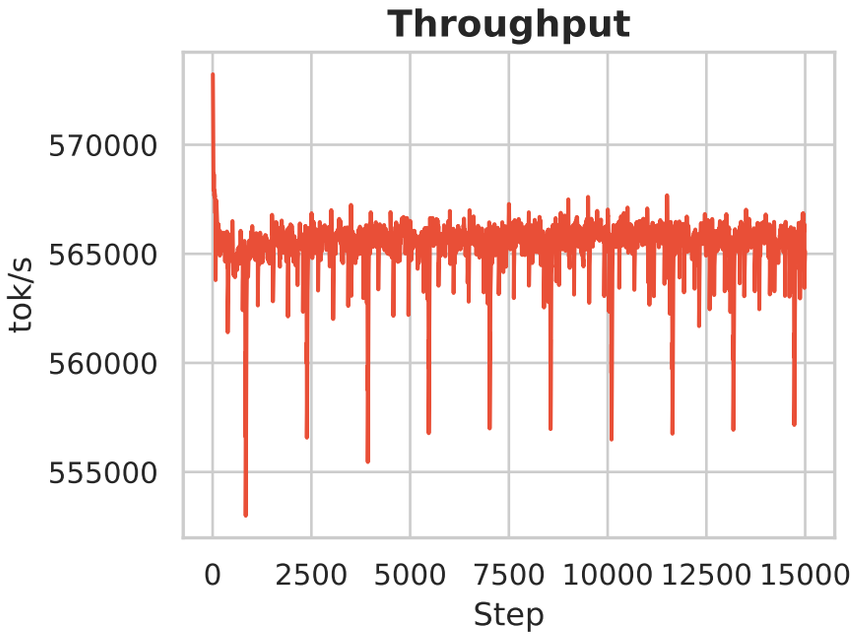
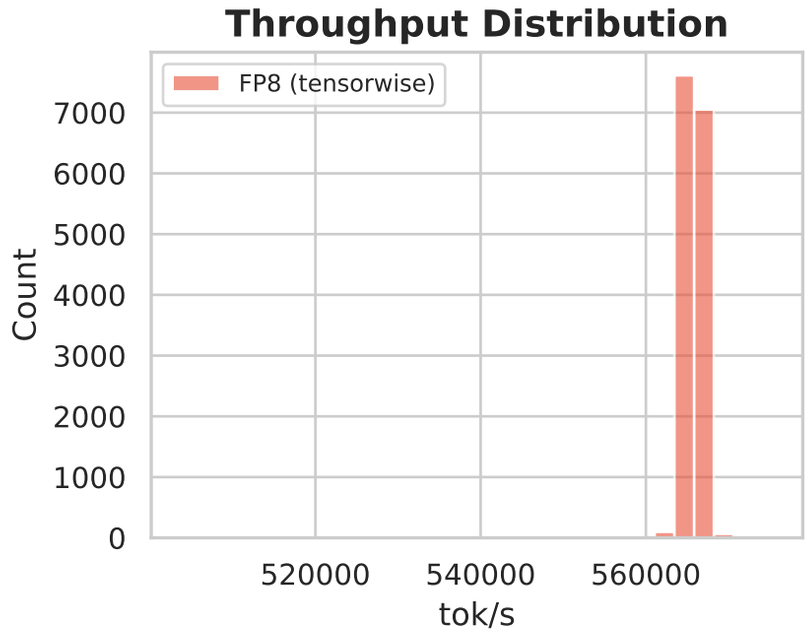
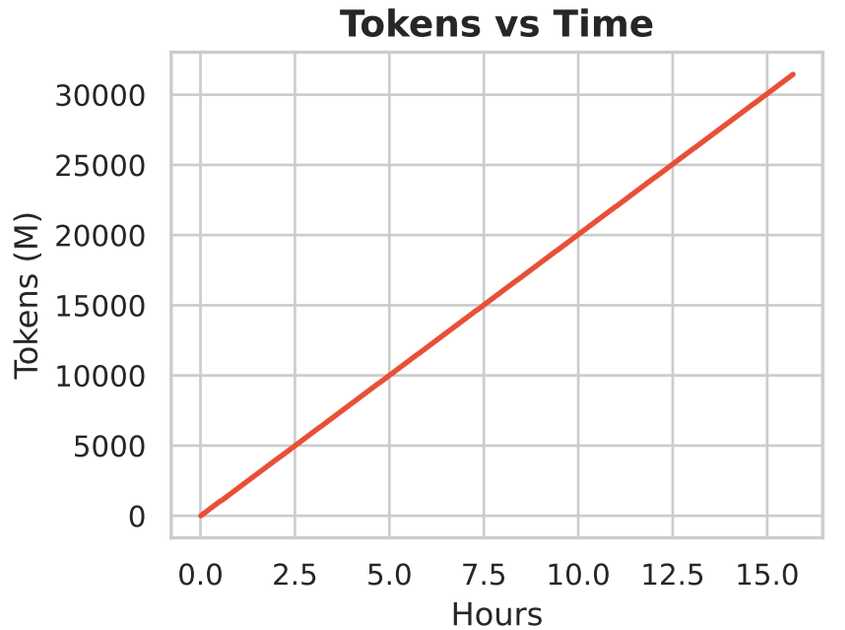
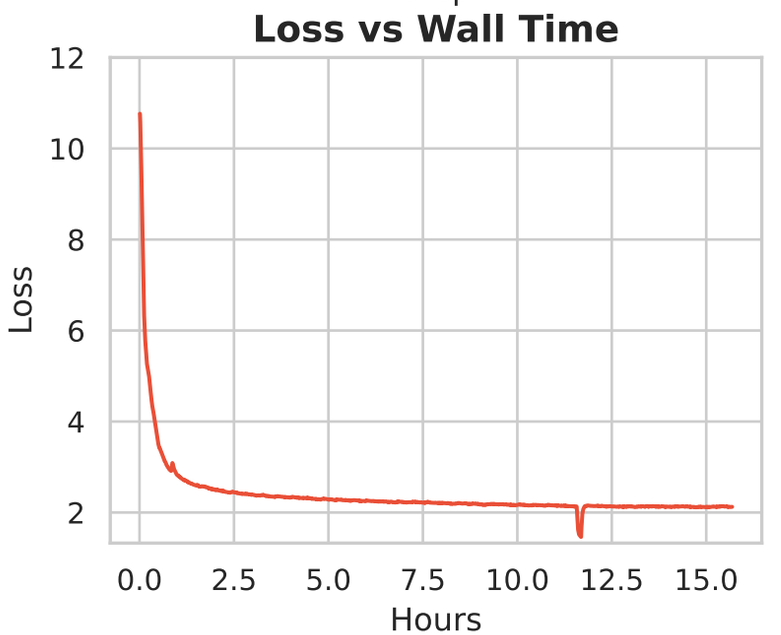
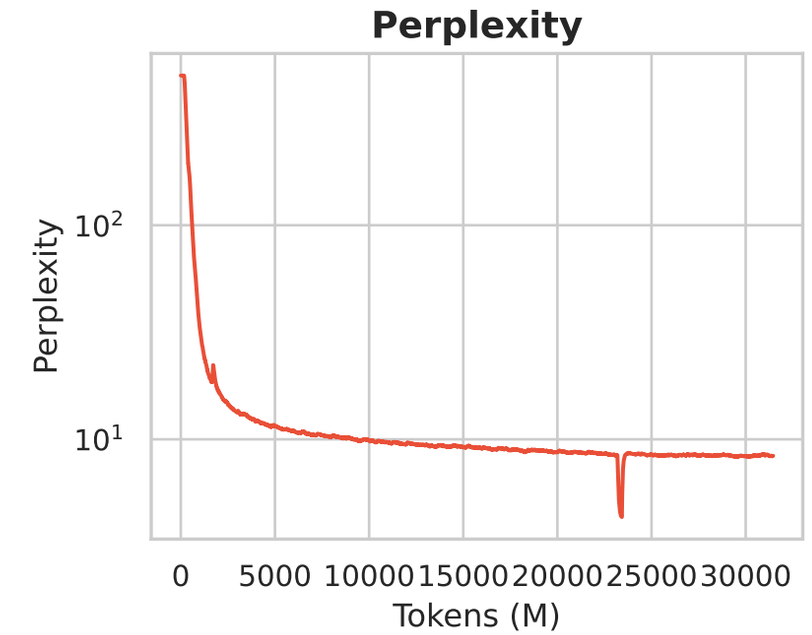
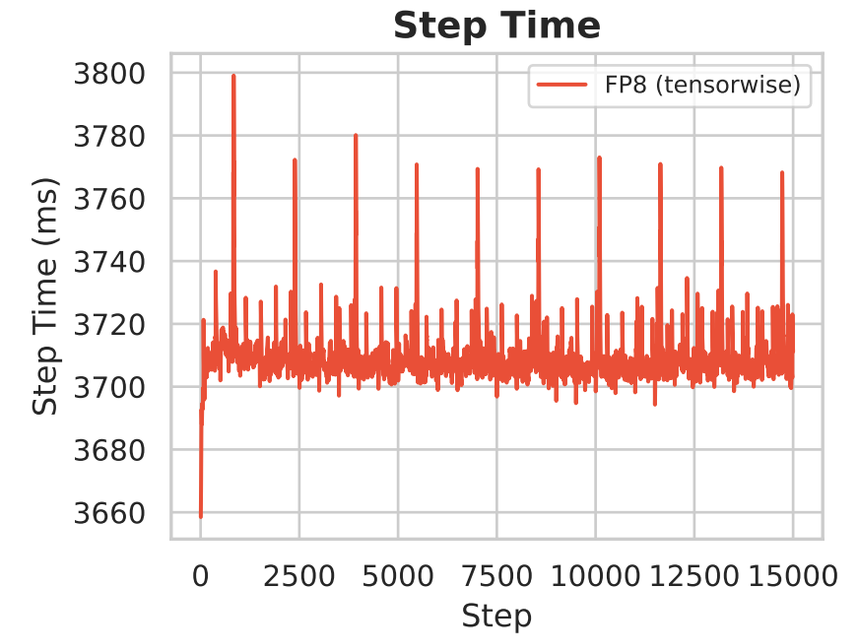
FP8 vs BF16 Training Comparison - 1.5B Parameter Model on NVIDIA GB10




```
Training Summary
==============================

FP8 (tensorwise)
  Final Loss: 2.128
  Final PPL: 8.4
  Tokens: 31457.3M
  Time: 15.69h
  Avg tok/s: 565518
```