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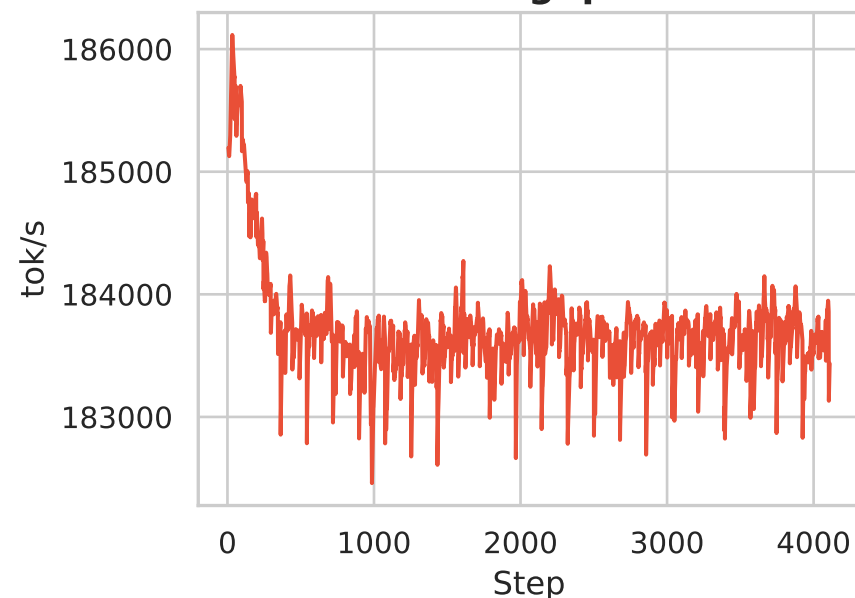
Training Loss
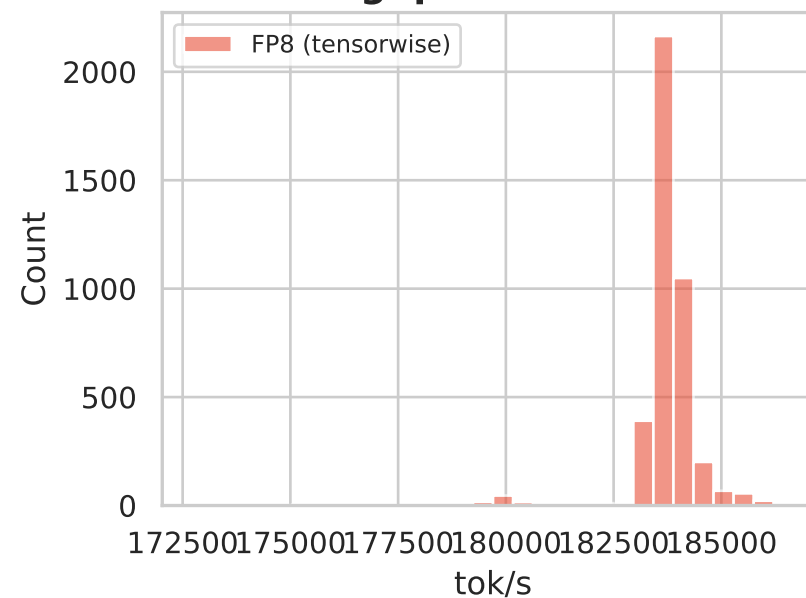


```
Training Summary
==============================

FP8 (tensorwise)
  Final Loss: 2.712
  Final PPL: 15.1
  Tokens: 8349.9M
  Time: 12.69h
  Avg tok/s: 183712
```
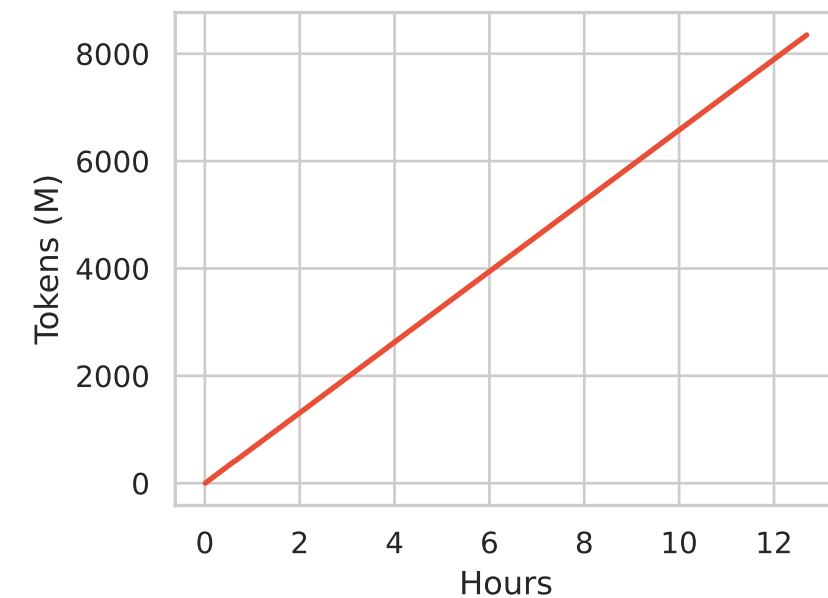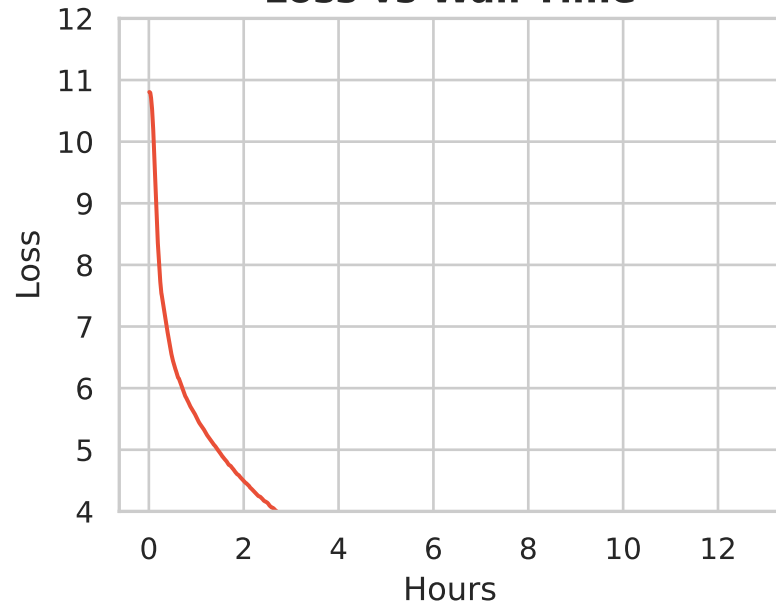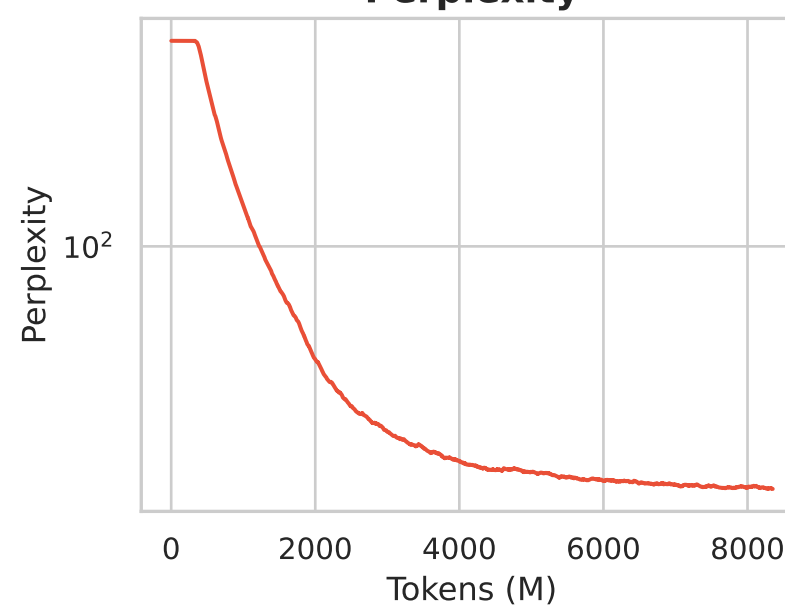
Throughput

Throughput Distribution

Tokens vs Time

Loss vs Wall Time

Perplexity

Step Time