The graph neural network (GNN) has become a popular research area for its state-of-the-art performance in many graph analysis tasks. Recently, various graph neural network libraries have emerged. They make the development of GNNs convenient, but their performance bottlenecks on large datasets are not well studied. In this work, we analyze the performance bottlenecks in GNN training and inference with GPUs empirically. A GNN layer can be decomposed into two parts: the vertex and the edge calculation parts. According to their computational complexity, we select four representative GNNs (GCN, GGNN, GAT, GaAN) for evaluation. We decompose their running time and memory usage, evaluate the effects of hyper-parameters and assess the efficiency of the sampling techniques. The experimental evaluation indicates that the edge-related calculation is the performance bottleneck for most GNNs, dominating the training/inference time and memory usage. The sampling techniques are essential for GNN training and inference on big graphs with GPUs, but their current implementation still has non-trivial overheads in sampling and data transferring.