BREAKING GLOBAL BUSINESS BARRIERS

Real-time translation plays an increasingly important role in the growth of global e-commerce. The Alibaba Group services 8.3 billion translation requests daily, enabling international business trade for small to medium-sized enterprises. Alibaba uses neural machine translation (NMT) to fill the gap between machine and human translation, which improves the translation quality significantly, but drives up latency and computation cost. To remedy this, Alibaba deployed NVIDIA Tesla GPUs to accelerate online NMT-based services, achieving a 3X increase in number of requests serviced, and cutting latency by a factor of three.