Parameter servers for machine learning do not scale?
Well, one size does not fit all.

**Motivation**
- Parameter Servers (PSs) facilitate distributed machine learning
- But existing PSs are inefficient for non-uniform access:
  - Skew: PSs inefficient because they manage all parameters identically
  - Sampling: PSs inefficient because sampling entails randomized access

**NuPS**
1. Supports multiple management techniques and picks a suitable one for each parameter:
   - Replication is efficient for hot spot parameters
   - Relocation is efficient for long tail parameters
2. Supports sampling directly via suitable primitives and sampling schemes that allow for a controlled quality—efficiency trade-off

**Results**
1. NuPS outperformed existing PSs by up to one order of magnitude across multiple ML tasks
2. NuPS provided up to linear scalability
3. Most efficient was to replicate a small set of hotspot parameters and relocate all others

Parameter server performance for training large knowledge graph embeddings on an 8-node cluster.