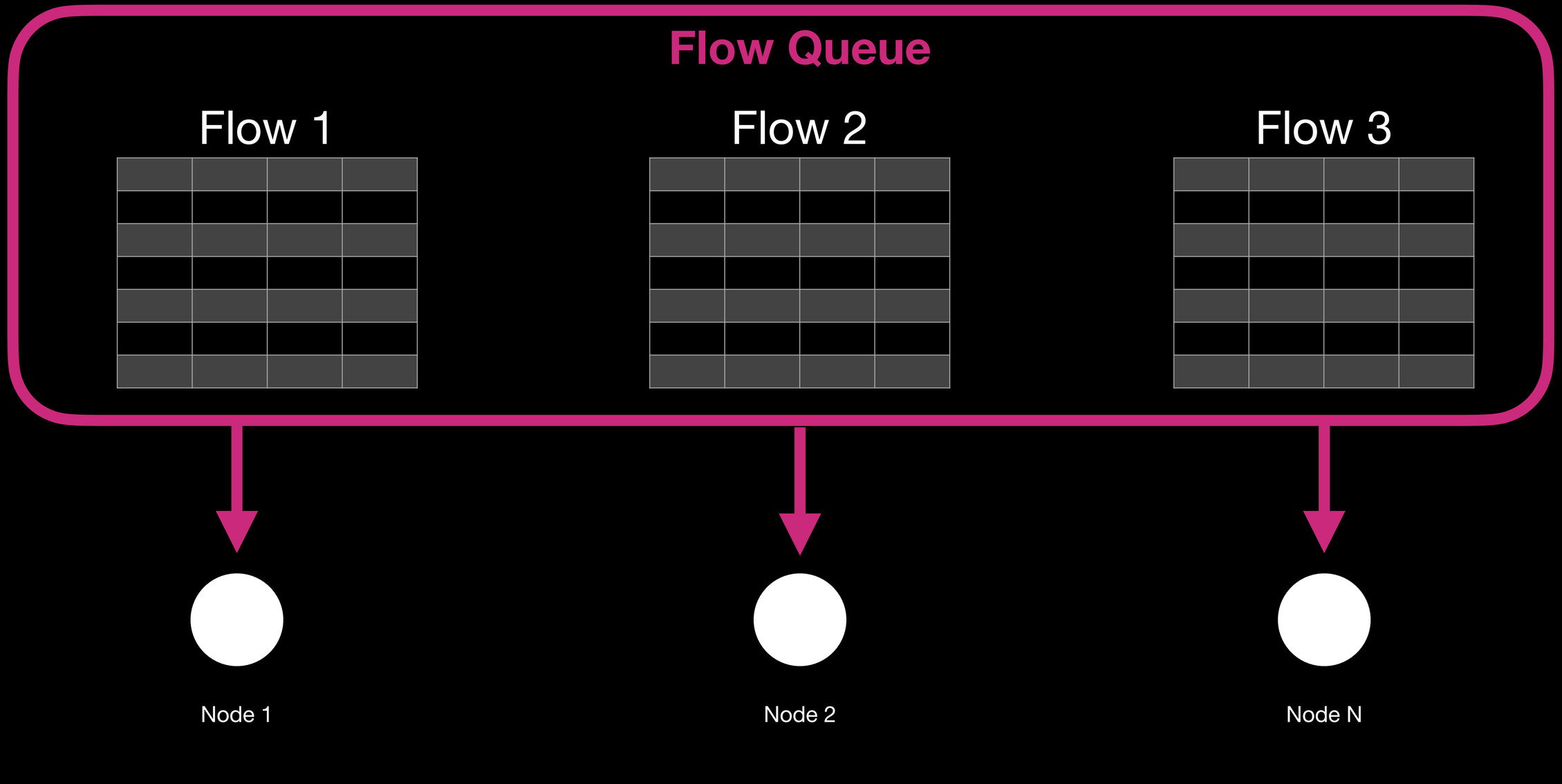
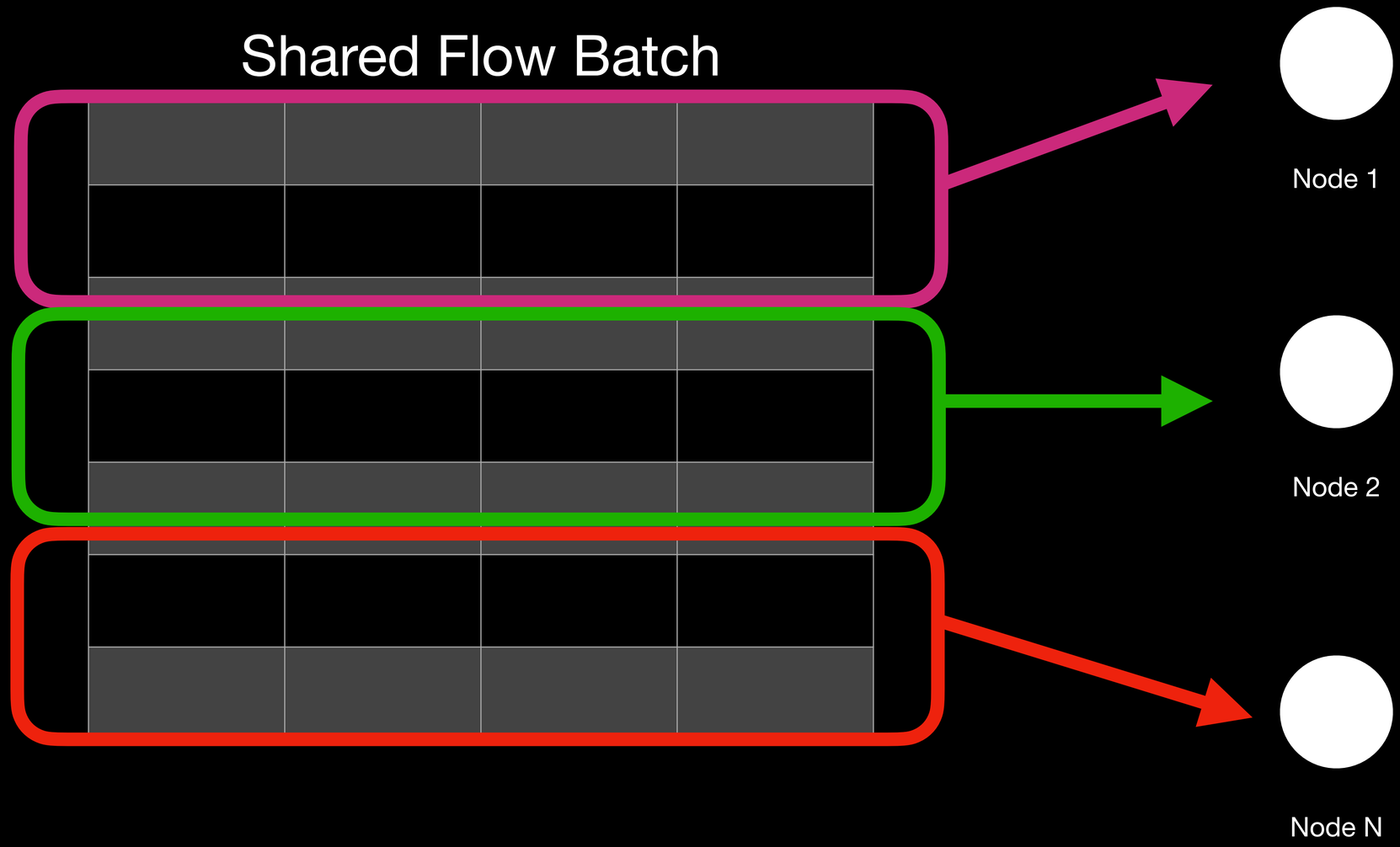
Data Delegation - Options

- **Option 1:**
  - Collect flows in increments and each node in the cluster gets handed a batch - our current approach.
- **Option 2:**
  - Collect flows into a larger batch, and have it distributed across the cluster with batch learning (available as an implicitly parallel operation in Ray).

Data Delegation - Option 1



Data Delegation - Option 2



From Zero (offline) Deployment Pipeline

- Step 1: Train all known nodes in a **federated** manner.
  - 1.5: Consider “best” model; e.g., accuracy, performance.
- Step 2: Begin IDS System
  - Consider data delegation strategy (collaboration).
- Step 3: Scaling
  - New nodes can join asynchronously and be served the model - already trained in step 1.
  - Now these nodes obtain their own flow batches and can predict using the model.
  - Many nodes can cause many aggregation requests (master node needs to be reasonably resource-capable).