MapReduce-Farthest-First Traversal

Let P be a set of N points (N large!) from a metric space (M, d), and let $k > 1$ be an integer.

Algorithm MR-Farthest-First Traversal

- **Round 1:**
  - Map Phase: Partition P arbitrarily into ℓ subsets of equal size $P_1, P_2, \ldots, P_\ell$.
  - Reduce Phase: for every $i \in [1, \ell]$ separately, run Farthest-First Traversal on P_i to determine a set $T_i \subseteq P_i$ of k centers.
- **Round 2:**
  - Map Phase: empty.
  - Reduce Phase: gather the coreset $T = \cup_{i=1}^{\ell} T_i$ (of size $\ell \cdot k$) and run, using a single reducer, Farthest-First Traversal on T to determine a set $S = \{c_1, c_2, \ldots, c_k\}$ of k centers, and return S as output.