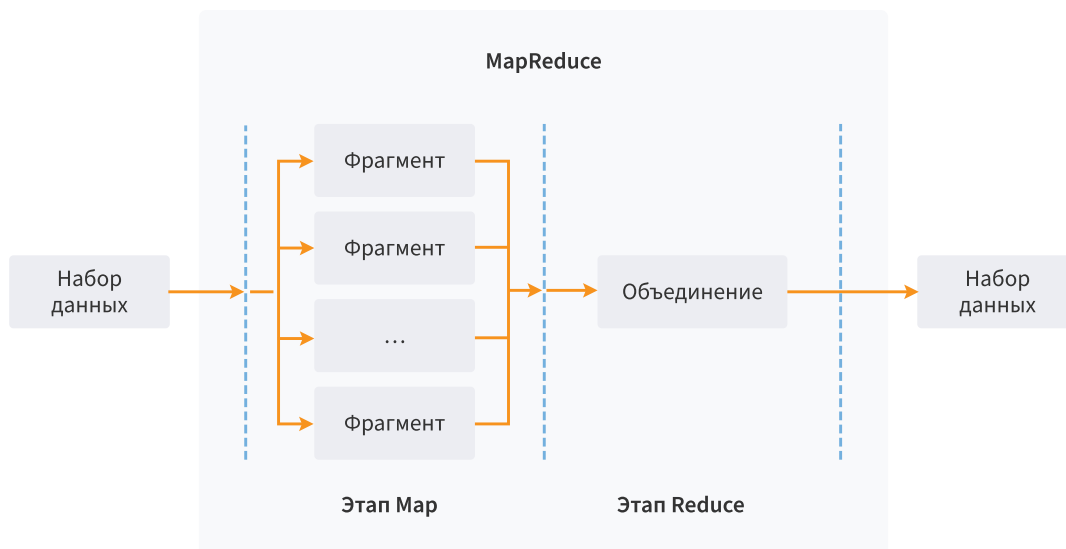


Модель MapReduce (MapReduce model)

MapReduce — это программная модель и комплекс связанных с ней процедур генерирования и обработки больших массивов данных с помощью распределенных вычислений на большом числе узлов. Модель представлена компанией Google, которая до 2014 года использовала MapReduce в качестве основной модели обработки больших данных.

Работа программы, основанной на модели MapReduce, состоит из двух основных шагов:

- **map** (отображение) — выполняется предварительная обработка данных на компьютере, называемом главным узлом (master-node), где они фильтруются, сортируются, разделяются на части и распределяются по рабочим узлам;
- **reduce** (сокращение) — главный узел получает от рабочих узлов обработанные данные и формирует из них результат, т.е. решение поставленной задачи.



Программы, использующие MapReduce, автоматически распараллеливаются и исполняются на кластере, состоящем из множества связанных между собой вычислительных узлов. Исполнительная система разделяет входные данные на части, планирует исполнение программы на наборе узлов, обрабатывает свои и управляет необходимыми коммуникациями между узлами.

Модель MapReduce применима в широкой области задач, включая распределенный поиск и сортировку данных, обработку статистики логов сети, кластеризацию документов, машинное обучение и машинный перевод.

