Twister – Iterative Map Reduce

Motivation for Twister

- Map Reduce framework won’t work for iterative tasks.
- It requires loading of large data set for each iteration which in turn splits data into 64 MB or 128 MB for each map worker.
- It is time consuming and we require a way to store results of mapper and use them in subsequent iterations,

Examples of Iterative Algorithms

- K- Means Clustering.
- Page Rank Algorithm.
- Matrix Multiplication.
- Multi-Dimensional Scaling.
- Breadth-First Search.

Twister extends Map Reduce to Iterative algorithms

Twister – A runtime for Iterative Map Reduce

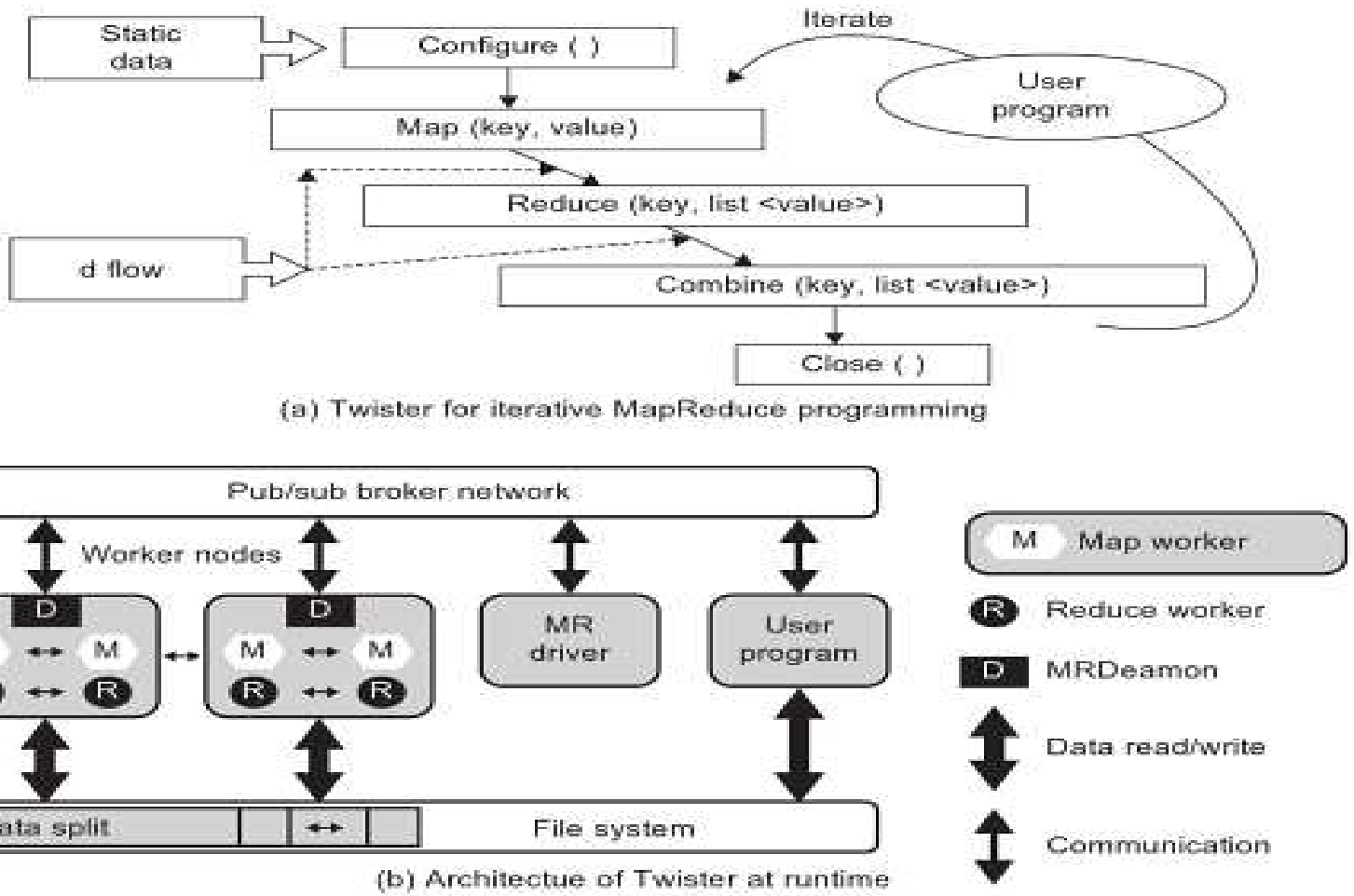
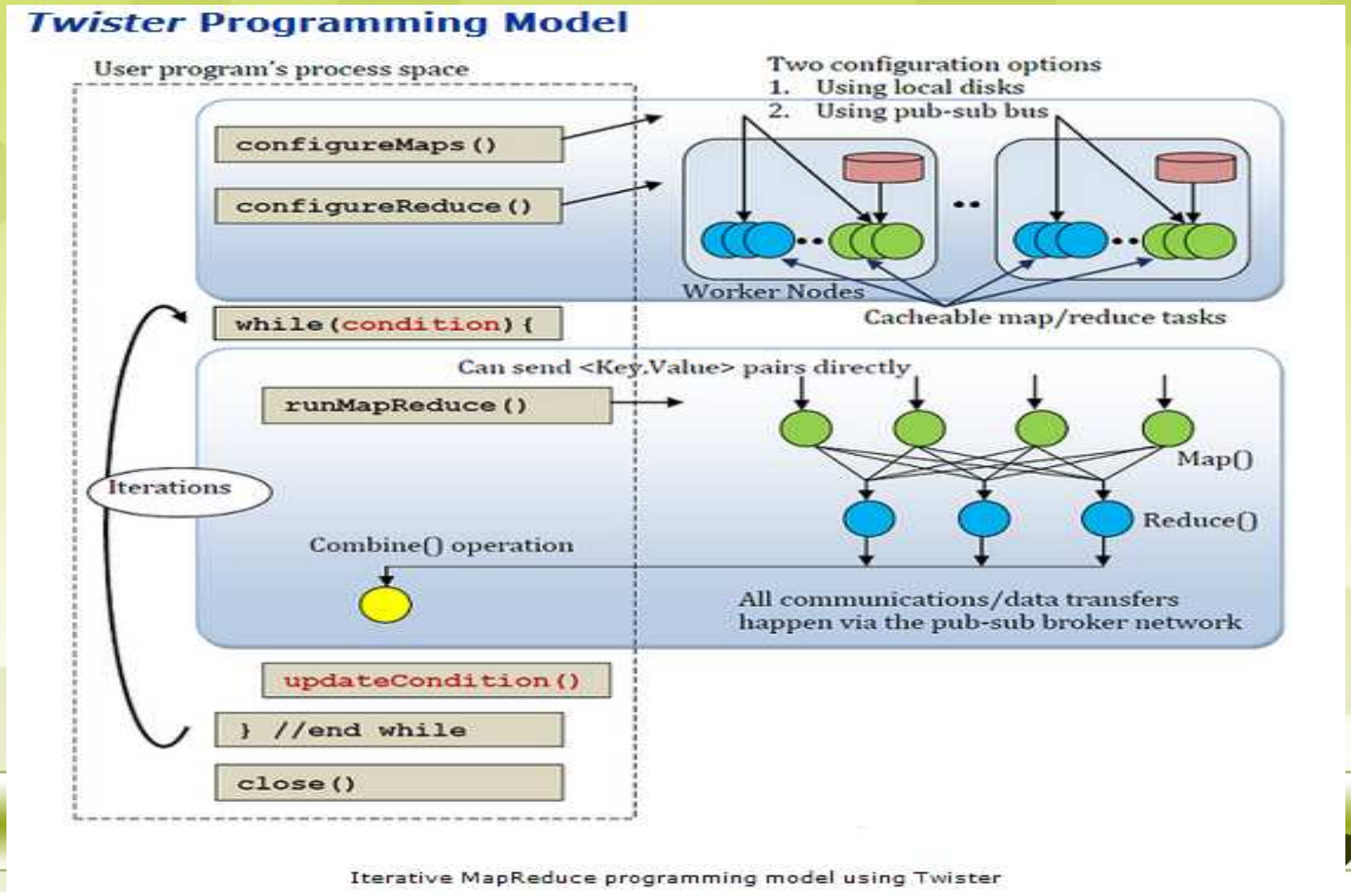


FIGURE 6.7

Twister: An iterative MapReduce programming paradigm for repeated MapReduce executions.

Twister – A runtime for Iterative Map Reduce



Iterative MapReduce programming model using Twister

Twister – A runtime for Iterative Map Reduce

- Distributed data access.
- In-memory Map Reduce
- Distinction on static data and variable data. (data flow vs. delta flow)
- Cache-able Map/Reduce tasks (long running tasks).
- Combine operation
- Supports fast intermediate data transfer

Twister – A runtime for Iterative Map Reduce

- Using local disks (only for maps)
- Using Publish / Subscribe (pub/sub) bus – communication / data transfer via pub/sub broker network.
- One broker in pub/Sub broker network can serve multiple Twister daemons.

SPARK & Mahout on top of Hadoop

- SPARK is an Apache framework which allows results of map to be stored in-memory and use them for subsequent iterations.
- The execution time of a task greatly reduces with SPARK.
- The following are used to run machine learning algorithms with Big data on Hadoop.
  - SPARK MLlib (Machine Learning Library)
  - Mahout

Open Challenges in Map Reduce

- Locality
- Task Granularity
- Dealing with Stragglers (Stragglers are slow working nodes which will reduce performance of entire MapReduce)
- Saving Bandwidth
- Handling of bad records.

Other Parallel Programming Models

- Dryad
- DryadLINQ
  - Both are Microsoft-based.

Thank You.