



How MapReduce Works

School of Information Studies
Syracuse University

MapReduce

A programming model for large-scale distributed data processing.

Foundations in functional programming, LISP.

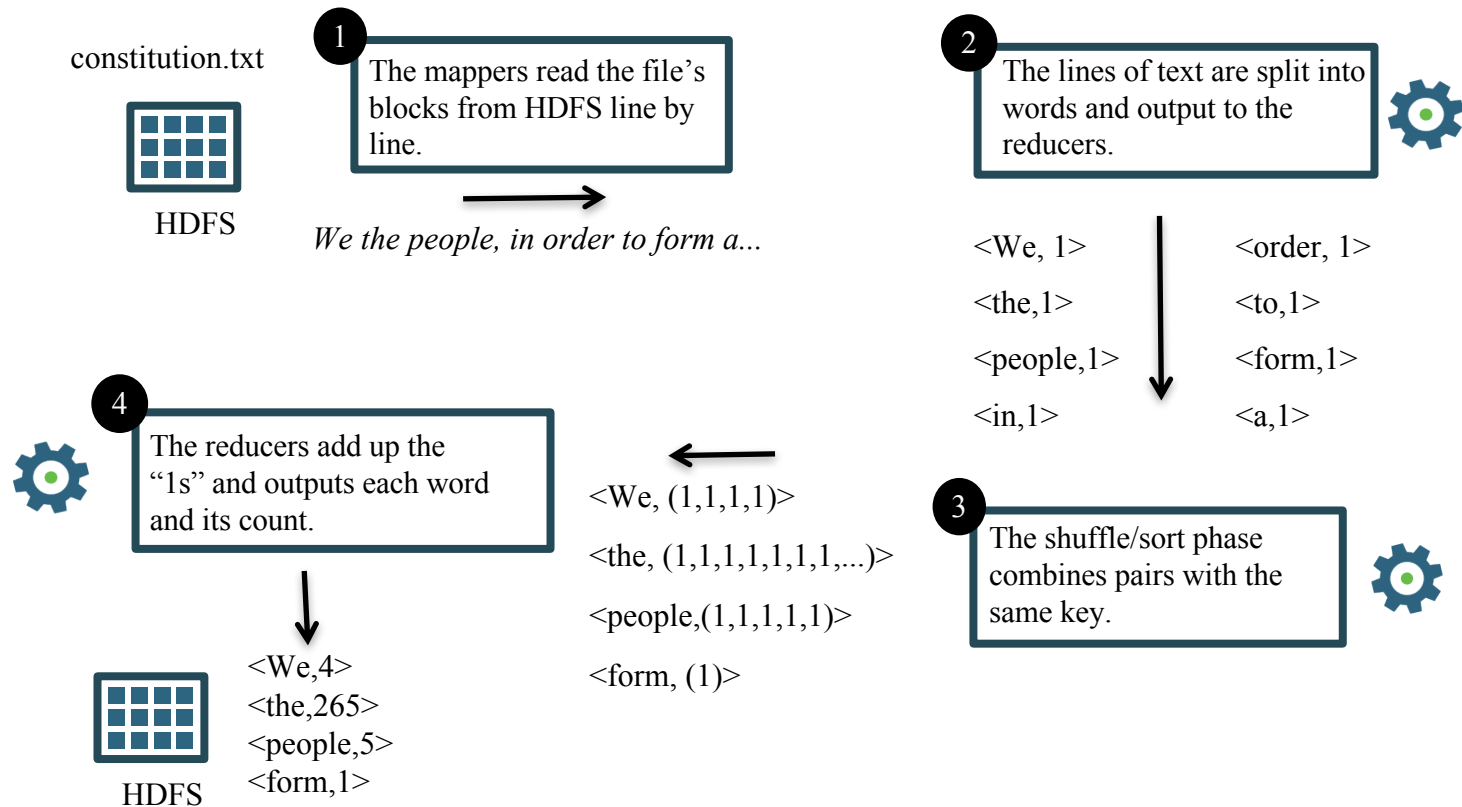
- Map → Apply a transformation to a dataset.
- Shuffle → Transfer output from mapper to reducer nodes.
- Reduce → Aggregate items into a single result.
- Combine → Output of reducer nodes into single output.

Hadoop 1.0 had only MapReduce.

Hadoop 2.0, MapReduce programs use YARN.



WordCount in MapReduce



MapReduce: Total Orders by State

