I have implemented the following items for optimisation.

* Prefix filtering to minimise the number of items emitted from the mappers.
* Hash map data structure to reduce the number of items emitted from the mappers.
* Proper delimiters to minimise the number of times to ‘split’ the input strings.