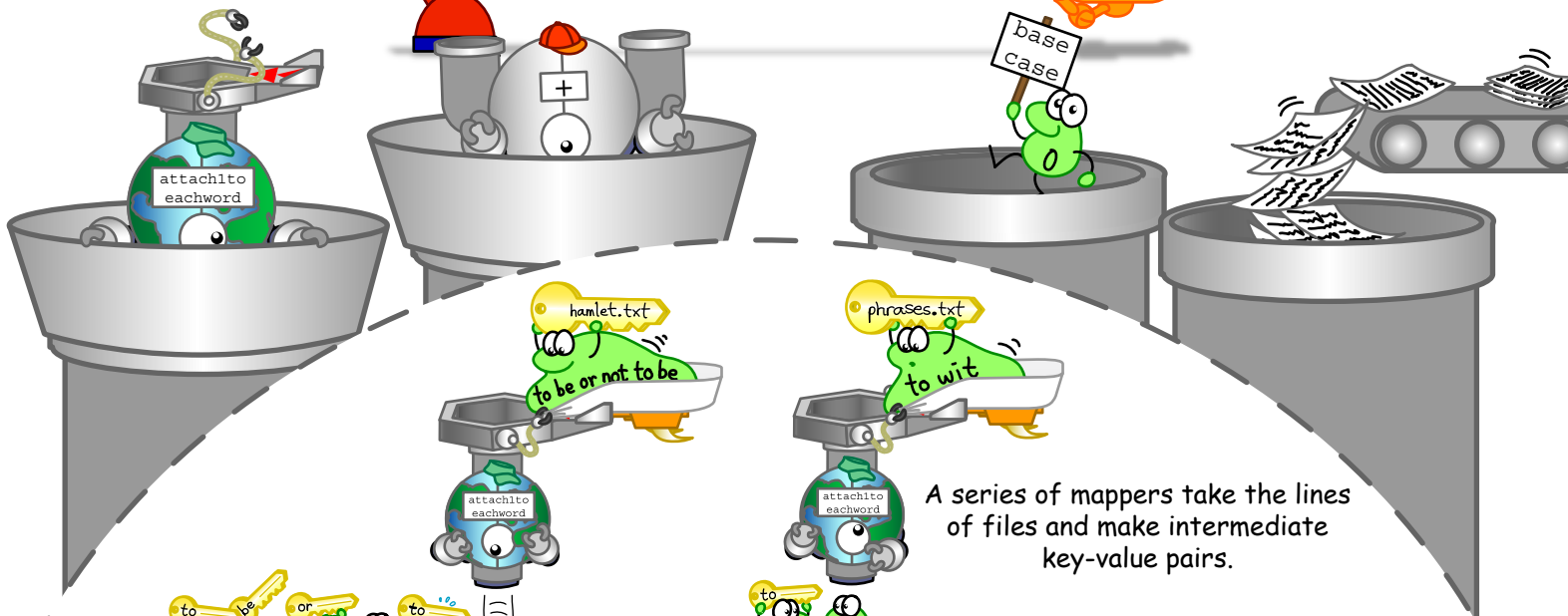


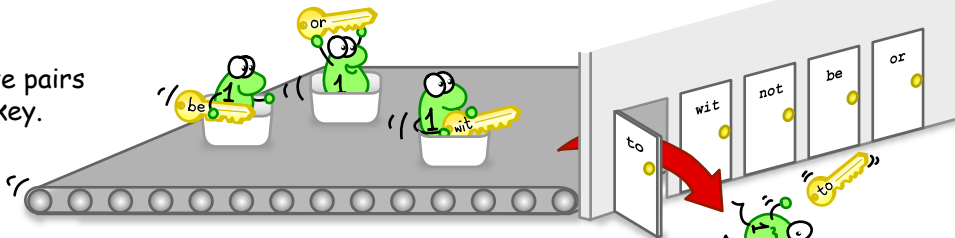
# MapReduce

An Example: Distributed Word Count



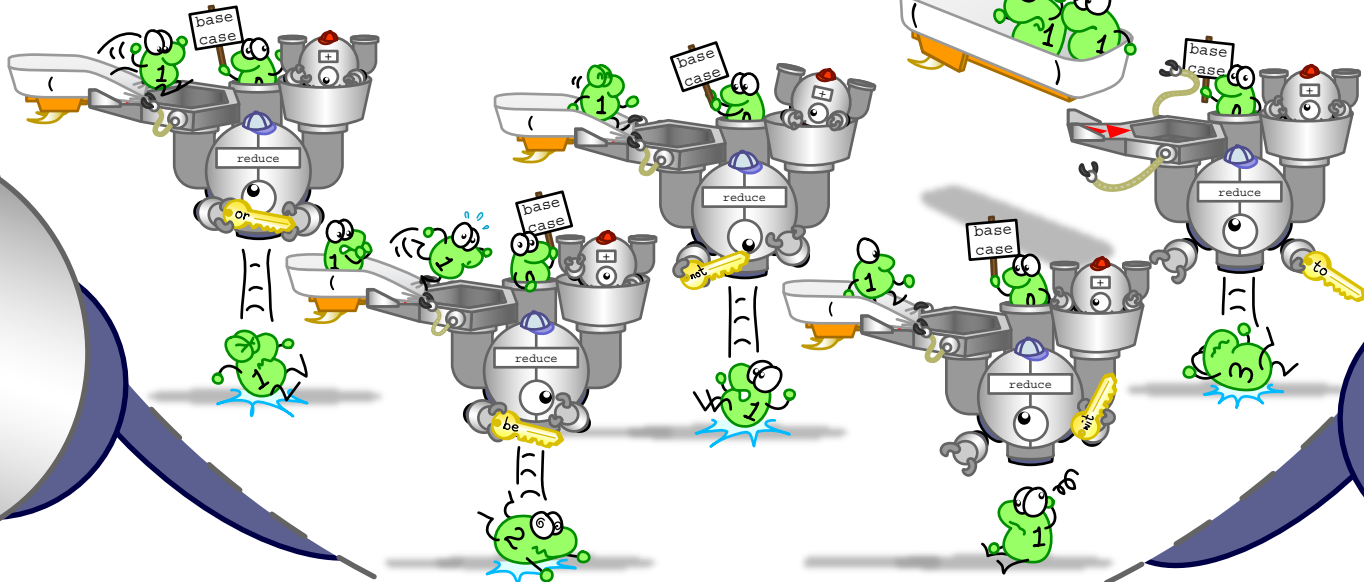
A series of mappers take the lines of files and make intermediate key-value pairs.

The intermediate pairs are grouped by key.



Using the input reducer, reduce processes combine the grouped data values into a single value.

Keep in mind that we've drawn the reduce processes with keys to make this step clearer, but in reality they don't know the keys!



The output emerges as a stream in an arbitrary order. Users can query this stream to find the count of a specific word.

