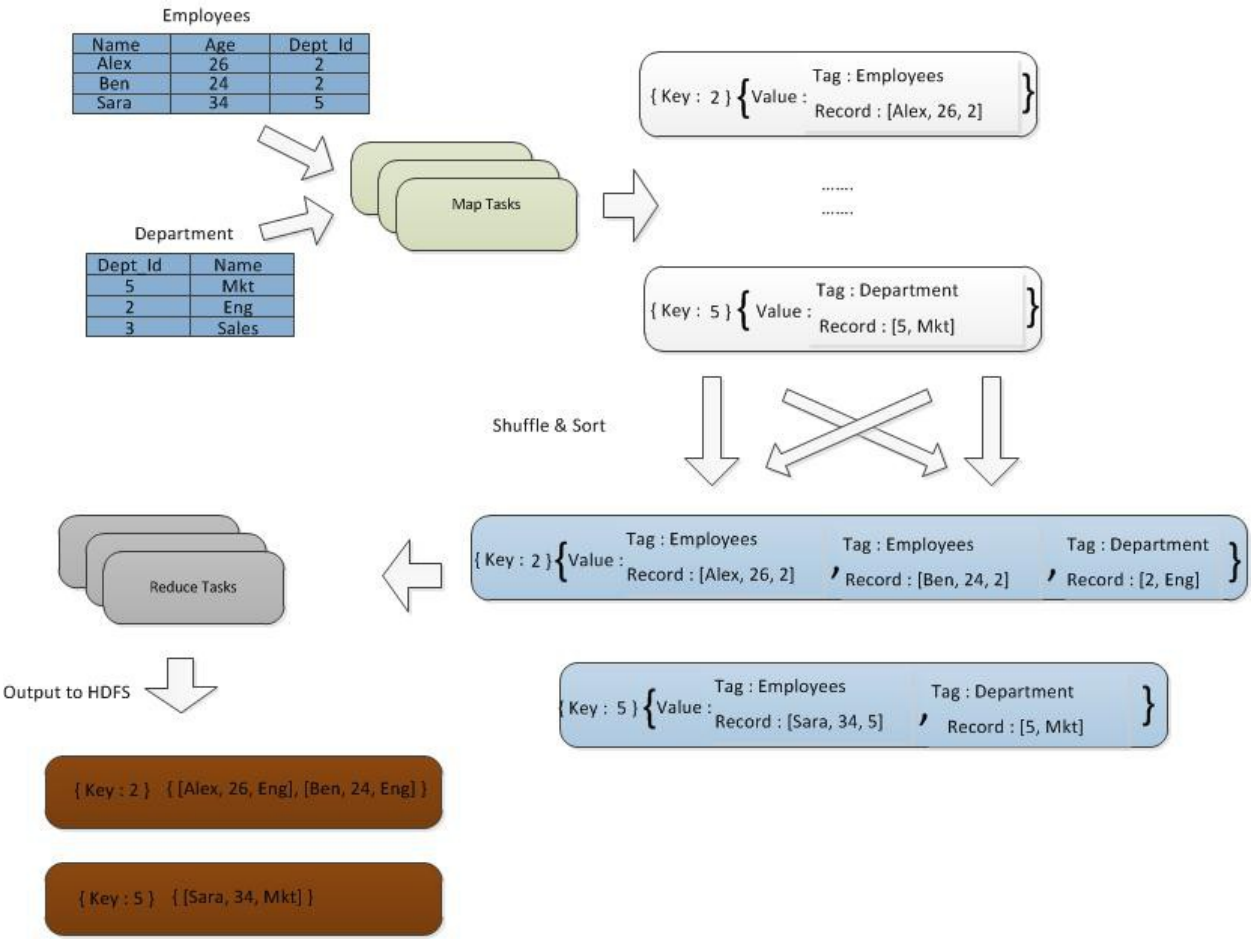
DDS ASSIGNMENT 4
ASU ID: 1211112666
NAME: ARCHANA RAMANATHAN SESHAKRISHNAN
EXECUTION COMMANDS:

hadoop jar <jarfile path> equijoin <hdfs_input_file> <hdfs_output_file_name>

To view the results:
hadoop fs -cat <hdfs_output_file_name>/part-r-00000

OVERALL JOIN LOGIC FOR MAP-REDUCE JOBS:

Reference: https://chamibuddhika.wordpress.com/2012/02/26/joins-with-map-reduce/



MAPPER LOGIC:

It divides the given input based on the join column value as key.

The MyMapper takes an input of object type and text type (input file) and gives Text as key and Text as value as output. The key is chosen to be Text even though the joining column has been given to be numeric because it has to accommodate integer, float and real values. Mapper output is given as input to Reducer.

DDS ASSIGNMENT 4
ASU ID: 1211112666
NAME: ARCHANA RAMANATHAN SESHAKRISHNAN
SHUFFLING AND SORTING:

It is done internally by the system itself.

REDUCER LOGIC:

Based on the set of column values given from MyMapper only the values from the previous are given as keys. This is because we need to only display the values based on keys from the mapper. Hence to adjust the output display the values from Mapper is given as key and value is NullWritable.

It is printed in a way desired by simple string concatenation and trimming.

MAIN FUNCTION:

Reference: https://hadoop.apache.org/docs/stable/hadoop-mapreduce-client/hadoop-mapreduce-client-core/MapReduceTutorial.html

The main function starts by creating Configuration and Job object. It is continued by setting various attributes of job object and Input and output paths.