|  |  |  |  |
| --- | --- | --- | --- |
| **Input-Splits** | {cat, bat, hat}  {mat, sat, rat}  {pat, sat, bat} | {pat, cat, mat}  {rat, hat, cat}  {pat, cat, mat} | {sat, bat, pat}  {rat, sat, mat}  {sat, bat, hat} |
| **Mapper Output** |  |  |  |
| **Shuffle and Sort** | **Assumuptions:**  **cat is 25, bat = 27, hat = 29, mat = 31, sat = 33, rat = 35, pat = 37**  **The number of reducers = 3**  **Mapper i, Reducer i on the same datanode for all i.** | | |
| **Reducer Input** |  |  |  |
| **Reducer Output** |  |  |  |

Name \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Class Number (**1 to 25**)\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Illustrating WordCount**

Q1. How many key-value pairs are shuffled across the network?

Q2. How many key-value pairs are shuffled across the network if we use In-Mapper Combining?

Q3. Assuming major cost factor is network traffic, what we accomplished?