# **CS 335:** Introduction to Large Language Models

*Habib University*

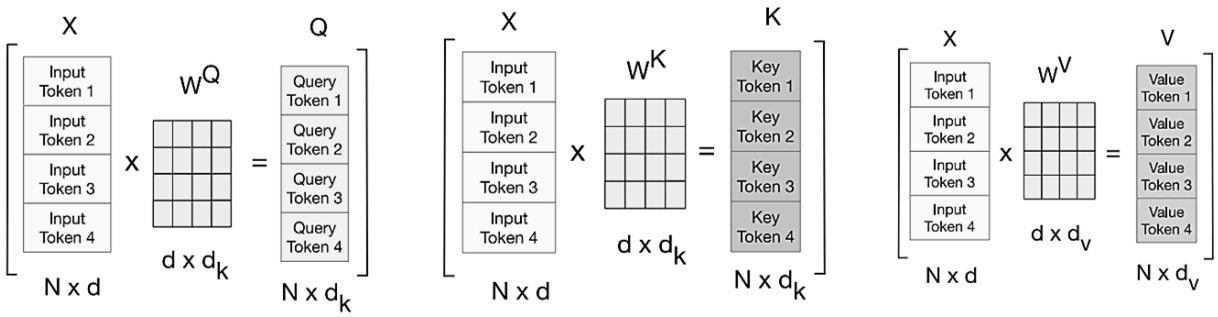
## Activity Sheet 04

Name: _____ ID: _____

**Question 01: Self Attention**

Computing self-attention is comprised of two stages:

In the first stage, we compute the matrices $\mathbf{Q}$,$\mathbf{K}$,and $\mathbf{V}$




Self-attention ($A$) is then computed as follows:

$$\textbf{Self–Attention (A)} = \left( \text{softmax}\left( \frac{\mathbf{QK^T}}{\sqrt{d_k}} \right) \right) \mathbf{V}$$

Given that $\mathbf{X}$ is as follows:

$$\mathbf{X} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{pmatrix}$$

where each row of X represents a word in the sequence.

The weight matrices $\mathbf{W}^Q$, $\mathbf{W}^K$, and $\mathbf{W}^V$ are:

$$\mathbf{W}^Q = \begin{bmatrix} 0.1 & 0.2 & 0.3 \\ 0.4 & 0.5 & 0.6 \\ 0.7 & 0.8 & 0.9 \\ 1.0 & 1.1 & 1.2 \end{bmatrix}, \quad \mathbf{W}^K = \begin{bmatrix} 0.3 & 0.2 & 0.1 \\ 0.6 & 0.5 & 0.4 \\ 0.9 & 0.8 & 0.7 \\ 1.2 & 1.1 & 1.0 \end{bmatrix}, \quad \mathbf{W}^V = \begin{bmatrix} 0.7 & 0.8 & 0.9 \\ 0.4 & 0.5 & 0.6 \\ 0.1 & 0.2 & 0.3 \\ 1.3 & 1.4 & 4.5 \end{bmatrix}$$

Compute Self-Attention (**A**).