

**CSCI 381/780: Special Topics on Computer Science (Cloud Computing)**  
**Spring 2023**  
**Mid-term Exam**

---

1. Describe the benefits of virtualization for cloud computing.
2. Describe (at least five) features of cloud computing. Justify your answer with details.
3. What is PUE? How is PUE calculated? Describe the factors that may affect PUE.
4. Compare the concepts of region, availability zone, and data center.
5. Why is the container more lightweight than the VM?
6. Why is Kubernetes neither IaaS nor PaaS?
7. Why does MapReduce need to ensure that all key-value pairs with the same key are shuffled to the same reducer? How is it implemented in Hadoop?
8. Given a file of 10 GB saved in HDFS with 256 MB as its chunk size ( $1 \text{ GB} = 1024 \text{ MB}$ ), how many map tasks are there in a MapReduce job taking this file as its input? Assume that the this job performs word counting with 4 reduce tasks. If the size of this job's output is 1 GB and each word appears 50 times on average, what is the average amount of data that a reduce tasks receives from a map task?
9. Compare the difference between the transformation and the action in Spark.
10. Calculate  $1 + 2^3 + \dots + n^3$  in Spark.