

Cloud Computing and Big Data Analytics

Spring, 2022

Homework #4

1. **[8 pts]** What are main three categories of cloud service models? Please briefly introduce them (**6 pts**). Suppose the electricity is very expensive in Taiwan. Which category is not suitable for providing this service in Taiwan? Justify your answer (**2 pts**).

2. **[20 pts]** Google offers both Software as a Service (SaaS) and Platform as a Service (PaaS) Solutions in Cloud Computing. Assume you plan to build a startup company with flower image classification by a machine learning approach. Discuss how your company can use Google App Engine (**10 pts** with some detailed steps) and the company potential cost (**4 pts**). Discuss the advantages and disadvantages by comparing with other popular cloud service providers such as Amazon and Azure Cloud Service (**6 pts**).

3. **[7 pts]** Virtualization plays an important role in cloud computing. Please state the difference between Docker and Virtual Machine (**4 pts**). Provide an application that is suitable for using Docker but is not suitable for using Virtual Machine (**3 pts**).

4. **[25 pts]** Based on data used in Lab 2, now write a MapReduce code to find the *variance* of monetary value for each separate store.

5. **[40 pts]** Based on the data used in Lab 2, now we want to find the frequent 2-itemset in Chicago (minsup=30%). We first group the items in the same day as one transaction. For example, if we only have the following three transactions in Chicago on 2012-01-01:

2012-01-01	12:08	Chicago	Music	10.99	Amex
2012-01-01	12:10	Chicago	Books	11.99	Amex
2012-01-01	12:18	Chicago	Cameras	222.99	Amex

We will group them into one transaction.

{Music, Books, Cameras}

Please write a MapReduce code to find frequent 2-itemset by using Apriori Based Algorithm.

[Hint] <https://arxiv.org/pdf/1702.06284.pdf>