1. What are main three categories of cloud service models? Please briefly introduce them. Suppose the electricity is very expensive in Taiwan. Which category is not suitable for providing this service in Taiwan? Justify your answer.

Ans:

Three categories of cloud service models including: (1) IaaS - Provides users the capability to provision computing and storage resources, users can deploy OS and applications on the resource. (2) PaaS - Provides users the capability to develop and deploy application in the cloud using the development tools, APIs, software libraries and services provided by the cloud service provider. (3) SaaS - Provides users a complete software application or the user interface to the application itself. Since the electricity is expensive in Taiwan, IaaS is not suitable for providing service here because IaaS aims to provide physical computing and storage resources including servers and IT infrastructures, which are electricity consuming.

2. 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 (with some detailed steps) and the company potential cost. Discuss the advantages and disadvantages by comparing with other popular cloud service providers such as Amazon and Azure Cloud Service.

Ans: Build machine learning model on PaaS, when the training process completed, it can play as role of SaaS for target clients to use the service for their own purpose (e.g. input their own target dataset and get predictions). The company can use Google App Engine to build service with steps below: 1. Establish project on platform and choose the location of server you want to use. 2. Upload the website and code to the platform. 3. Start the service. Google App Engine allows each developer account can register 10 applications with a total of 500MB of storage space, which can accommodate an estimated 5 million web page views per month. Additional fees will be charged if usage is exceeded, which are considered as potential costs.

Amazon Web Services (AWS) is the world's most comprehensive and broadly adopted cloud platform and it provides a wide range of

services that most branded companies use. (e.g. Netflix, BBC, Facebook, Spotify, LinkedIn). However, its cost is higher compared to other platforms, and its focus on public clouds rather than private or hybrid clouds which means that interoperability with data centers is not a major task for AWS. Azure has an outstanding cloud infrastructure and is the main competitor to AWS. Azure is rapidly entering the market with enterprise deployments of windows or other Microsoft software, and its advantage is that it can interoperate with the data center you are running to achieve hybrid usage. Google Cloud Platform (GCP) is a late entrant to the cloud market and has a less well-funded background than other platforms, but it has deep technical expertise and significant advantages in deep learning and artificial intelligence, machine learning and data analytics.

3. Virtualization plays an important role in cloud computing. Please state the difference between Docker and Virtual Machine. Provide an application that is suitable for using Docker but is not suitable for using Virtual Machine.

Ans: The operating system support of Virtual Machine and Docker container is different as the description below: Each Virtual Machine has its guest operating system above the host operating system, which makes Virtual Machines very heavy. Instead, Docker containers share the host operating system, and is lightweight in comparison with Virtual Machine. Sharing the host operating system between the containers make them lighter and helps them to boot up in just a few seconds. Therefore, the overhead to manage the container system is lower than that of virtual machines.

The docker containers are suited for situations where you want to run multiple applications over a single operating system kernel. But if you have applications or servers that need to run on different operating system flavors, then virtual machines are required.

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

Ans:

①mapper: mapper4.png ②reducer: reducer4.png

③result: q4-1.png, q4-2.png, q4-3.png

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

Ans:

①mapper: mapper5.png

②reducer: reducer5-1.png, reducer5-2.png, reducer5-3.png

③result: q5-1.png, q5-2.png, q5-3.png, q5-4.png, q5-5.png