1ai) For clients to access the application, the cloud needs to configure the IP addresses and firewall settings. The cloud begins by assigning a public IP address to the container enabling clients to reach the application. Using NAT (Network Address Translation) as a middlebox, the cloud translates the external IP address and external port into a private IP address and port, enabling the communication between the cloud application and clients is established. The firewall settings are adjusted by the cloud platform to allow clients to establish a connection with the containers, which is done by enabling the firewall to allow specific traffic for specific ports. (Week 2- slide 55)

1aii) The messages exchanged between clients depend on the need of the client. If a client wants to access a certain website, then server sends back the IP address of the DNS associated with the website. Another example is that a client may send HTTP requests to a server when the client requires certain resources where the server responds if this is okay or rejected.

1b) An advantage is that PaaS is simpler to use than IaaS and is more abstract making it easier to use. PaaS would be used over IaaS due to the disadvantage of PaaS having less customisation. This is because it offers less tools, or abstract components in comparison to PaaS. (Week 1, slide 55)

1ci) The first technique is emulation which simulates the hardware platform of the guest VM as close as possible and is used for virtualising different architectures. However, a drawback is that it has a large performance cost. The second technique is paravirtualization where the guest VM is modified so that it is unaware of its virtualisation status. The third technique is hardware-assisted virtualisation. This provides almost optimal performance to guest virtual machines that have the same physical architecture such as Intel VT-x and VT-d instruction sets. Considering all these techniques, emulation offers the best performance since it very closely mimics the comprehensive hardware as though it is the real entity. Whereas hardware-assisted virtualisation uses the physical architecture abilities which makes it the cheapest. (week 1, slide 11)

1cii) Containerisation is the best technique. It is an operating system-level virtualisation where the images in the containers only contain libraries and application code which enable a fast start-up and easy replication. Thus, containerisation is suitable for dealing with several services that have low resources. It may have has worse isolation than emulation and hardware assisted virtualisation, but the efficiency in the allocation of resources makes it better. (Week 1, slide 12)

2a) The server does not need to know the IP and port beforehand since it waits for incoming connections from clients. Since clients have to establish the connection, the client needs to make the request to the correct IP address and port so that the server receives it. The server does not need to know this information as it awaits for a client to send a connection.

2b)

2c) An advantage of REST services is due to its property of being stateless meaning that the server does not store any information of the previous interactions offering clients complete isolation, whereas in RPC, servers and clients are not decoupled so cannot develop by themselves. An advantage of RPC is that in the case of scenarios where you require the system to stateful, less information would need to be provided in each transmission resulting in a smaller payload size compared to the REST where all the information must be included each time. Thus, this gives improved latency and improved performance of RPC over REST.

2d) So firstly the client’s DNS query is transmitted to a DNS resolver where the DNS resolver passes the query on the authoritative DNS server and CDN authoritative DNS server. Then the CDN authoritative DNS server picks which CDN cluster to pass the request to i.e. CDN cluster 1 or CDN cluster 2. This decision is sent to the client through the DNS system where the client makes a HTTP request to the chosen CDN cluster i.e. CDN cluster 1. (Week 2, Slide 47)

2e) What are the security implications of first encrypting with your private key and then with your public one?

By carrying this out, this will provide confidentiality and non-repudation.

3a) Linearisability is when the outcome of an operation becomes visible to all nodes in the system meaning that if any follower replies to a replication request that it is okay, then all of the followers can see ensuring consistency. This gives the illusion that this process happened instantly. Non-linearisable is when the observer receives latency if the replication request is okay from the nodes. An example can be if you are depositing money into an account, if you deposit money in twice and then wish to withdraw an amount that is greater than the first deposit but less than the second, then using linearizability, if one node replied that you have enough money to withdraw then the action will be completed since the rest of the nodes will agree.

3b)

3c) Hashing passwords

Hashing is the process of where you apply a one-way function to some data where you receive a hashed string where the purpose of doing this is that it is irreversible, making it extremely difficult for an unauthorised entity to be able to make sense of this data. Thus hashing passwords is good because passwords to your email service or computer contains sensitive and personal information and by hashing passwords, it improves your security. (Week 8, slide 25)

3d) Asymmetric encryption is slower than symmetric because both the sender and receiver have their own public and private keys. In symmetric, the same key can be used to encrypt and decrypt whereas in asymmetric different keys are used.

3e) SLA defines the services provided by the cloud service and the metrics they promise to deliver to the consumer according to service metrics. QoS measures these metrics and checks if they agree with what was stated in SLA. Thus it makes sure that the resources are up to standard for the consumer to use.

3f) Vertical scaling would be used as you don’t have to split workload. You can upgrade the machine capabilities such as RAM and CPU enabling more machine learning workloads to be executed faster. You cannot split the data using horizontal scaling as having many more machines doesn’t suit the nature of what is best for the machine learning workloads, if this was done it would be less effective since the best performance is not achieved if the data in memory is split across a network. Some machine learning algorithms take a long time so a better CPU will help more.

4ai) Load balancing,

4aii) Micro-services architecture

4aiii) Live machine migration infrastructure may be used when there is heavy traffic such as when there is some worldwide entertainment occurring such as the world cup.

4b) We use analytical tools to measure and monitor the resources so when high demand approaches more resources can be allocated. The analytical tools can also have features where they use information from other real-world factors that can affect the demand. By doing this, you have better utilisation and performance for resources.

4c)

**Functionality:** Check status

GET

Path: /booking/{roomNo}

Get a response of the status of room is {free/filled}

**Functionality:** Request to book free room

POST

Path: /rooms/{roomNo}/bookings

Body: You are booked in the room from X until Y

Response of the booking is confirmed with the timings

**Functionality:** Cancel a booking

DELETE

Path: /rooms/{roomNo}/bookings/bookingNo

Receiving a cancellation reponse

4d) Request-driven can handle more requests so can be scalable by adding more servers. However, message-driven has better scalability overall as more queues can be added. By adding more queues you are able to facilitate for more messages.