Part A:

Q1: An IDS is a system that monitors and analyzes data to detect intrusions in a system or network. The major methods for detecting intrusions are signature-based, anomaly-based and hybrid methods which uses both. It is possible to implement an IDS on the 10% subset of the KDD cup dataset, with the same workflow as described in the paper. There are 3 main steps in the workflow. Firstly, the categorial data is converted to numerical data and the dataset is standardized to improve the efficiency at classification. This step is called data preprocessing. Then, the authors used ChiSqSelector to reduce the features in the dataset, which is called feature selection. The optimum is determined by model evaluation on different numbers of features. After feature selection, the svm classifier is used to determine whether it’s normal or malicious.
Q7: The first chart, distribution of duration, is a histogram of the duration. The second chart, the relationship between source bytes and destination bytes, is a scattered plot of the destination bytes and source bytes. The thirsd chart, distribution of labels, is a bar chart of the label counts.

Q8: Here, Logistic Regression is selected because it’s simple to implement and efficient for binary classification. In the model, we classify the connections as normal or attack. The accuracy of the model is ~85%. The confusion matrix shows the model predicts normal class pretty well, but faces challenges identifying attack ones.

Part B:

1.1 False. For PaaS, it only provides services that assist development, like Azure DevOps.

1.2 True. Because the infrastructure is managed by the cloud providers.

2. d. For a, dynamic schema is associated with NoSQL. For b, K-V type are typical NoSQL DB. For c, for the blobs, it’s better to use object storage services like S3 or Azure Blob Storage. For d, consistency is a part of ACID which is guaranteed by RDBs.

3. d. For a, b, c, the properties mentioned are managed by the providers.

4.1 False. Because continuously integrating public cloud services is possible.

4.2 True. Using a public cloud could add scalability and reliability.

4.3 False. Access is controlled by authentication and authorization, instead of restricting access to guest users only.

5 a. Fault Tolerance

b. Disaster Recovery

c. Dynamic Scalability

d. Low Latency