LINUX KURIOSITY

Devops: Data Science

[Interview-QA]

Learn in a Simple & Smart way

1. What is Devops ?

Ans: Devops is a process that combines software development with operations. It facilitates faster code deployment into production.

2. What is Data science ?

Ans: Data science is an interdisciplinary area that uses scientific methods , processes, algorithms and extracts insights from data.

3. What is training data ?

Ans: A Machine learning algorithm builds a model out of sample data, known as training data. The model is further used to make predictions of unknown data.

4. Give a one example where we can use machine learning algorithms.

Ans: Machine learning algorithms are used in applications of filtering spam emails.

5. What are the two tasks provides by machine learning to computers ?

Ans:

a. classification
b. predictions

6. What are the major steps involved in machine learning ?

Ans:

a. Data collection
b. Data munging
c. Feature engineering
d. model building
e. Testing and validating model
f. Running models

7. What is the first step of any data science project ?

Ans: The first part of any data science project is data collection.It enables organization to derive answers to specific questions. The data can be collected from multiple sources such as relational databases, NoSQL, spreadsheets , logfiles and so on.

8. What is Data Munging ?

Ans: Raw data is transformed into another format using techniques like filtering. This process of transformation is known as data munging.

9. What is the role of feature engineering in machine learning ?

Ans: Feature Engineering is the most crucial step in machine learning.

Some of the tasks carried in feature engineering are :

1. Identification of correlated data columns.
2. Creating new data columns from one or more existing columns.

10. What is the role of weights in machine learning ?

Ans: The training process involves initialization of random values, known as weights, for the chosen algorithm and then perform the predictions.

11. How the accuracy of built in model can be tested ?

Ans: The accuracy of built in model can be tested using a test data set.Models built must be accurate enough so that they can be used for decision making.

12. After models validation , what are the two ways can be used for prediction ?

Ans:

a. In the first case, the predict methods, associated with a model can be called directly from a script and obtain the result.

b. In the second case the model is moved into production and users interact with it over the internet .In this case, predict methods of a model are not directly accessible.

13. What are the three different environments in devops cycle ?

Ans:

1. Development
2. Staging
3. Production

14. What are the languages used by data scientists ?

Ans: Data Scientists use R/Python languages for building models and for consuming these models.

15. What are the two majorly followed approaches for deploying models in production ?

Ans:

a. Rewrite the whole code of building and accessing model in languages supported by stack environment.

b. Develop APIs for accessing models.

16. What are the basic uses of machine learning models ?

Ans:
a. Machine learning models are generally used for either classifications or regression problems

based on some input data.

17. What is continuous integration ?

Ans: The process of deploying models frequently into production is known as continous integration.

18. How continuous integration process can be automated ?

Ans: Continuous Integration process can be automated using a popular tool (jenkins).

19. What are the major practices to be followed for a deployment of machine learning model ?

Ans:

a. Using a version control system for models
b. Perform canary deployment
c. Secure models in production
d. Monitor performance of models in production.

20. How to keep track of different versions of a model ?

Ans: We can use version control systems.

21. What is canary deployment ?

Ans: In a canary deployment new changes to a model are exposed to a limited proportion of entire users.

22. How to secure ML models in production ?

Ans: Models can be secured using access controls like authentication and authorization.

23. What is drift ?

Ans: Drift means change in model input data.

24. What is web service ?

Ans: A web service is a software program available to users over the internet. A user invokes a web service by sending a web request and waits for its response.

25 . How a model will act if it is embedded in a program/API ?

Ans: It will act as a service. This service can be invoked through HTTP protocols. Also, service also acts as an abstraction to a machine learning model from users.

26. What is the role of a restful API ?

Ans: A restful API invokes a machine learning model functionality, executes it and captures the response.

27. What are the allowed HTTP methods ?

Ans:

a. GET
b. POST
c. PUT
d. DELETE

28. What is plumber ?

Ans: Plumber is used to quickly deploy models , built using R language , as services.
A REST API can be easily created by decorating a R function with special comments.

29. What is Flask ?

Ans: Flask is one of the python frameworks used to quickly deploy models, built using python language as services.

30. What is container ?

Ans: A container is a standard unit of software that packages up the code and all its dependencies, so the application run quickly and reliably from one computing environment to another.

A container consists of an application, all of its dependencies, libraries and other configuration files required for running it. Containers are different from virtual machines.

31. What is Application scaling ?

Ans: Application scaling refers to the alteration of computing resources required by the application , as load changes.

32. How applications can be scaled ?

Ans: Applications can be scaled vertically or horizontally.

a. Vertical scaling : It increases computing resources by adding more resources, directly to an existing server.

b. Horizontal scaling : It increases computing resources by adding more physical machine or servers.

33. What are the three main components of docker ?

Ans:

a. Docker engine
b. Docker client
c. Docker registry

34. What is docker engine ?

Ans: An application which runs docker images and also contains a command line utility.

35. What is Docker client ?

Ans: It contains tools for building and running docker images.It helps in interacting with other docker components.

36. What is docker registry ?

Ans: It stores docker images. Docker hub is a public registry, which can be used by anyone to share
and access public docker images.

37. What is docker image ?

Ans: A Docker image is a binary file, containing details of resources required to execute an application.

38. What is dockerfile ?

Ans: A Dockerfile is a configuration file that specifies the components to be included in an image.

39. Docker images are built from where ?

Ans: Docker images are built from docker files.

40. What is docker registry ?

Ans: Docker registry is a storage and delivery system for named docker images.

41. What is repository ?

Ans: A repository is just a collection of different versions of a single docker image ?

42. What are the two commands used to retrieve and store an image from a docker registry ?

Ans: Pull and push commands are used to retrieve and store an image from a docker registry.

43. How Docker images can be managed ?

Ans: Docker images can be managed using docker registries.

44. What is kubernetes ?

Ans: Kubernetes is an open source system for automating deployment, scaling and management of containerized applications.

45. What are the two types of nodes in kubernetes cluster ?

Ans: There are two types of nodes in kubernetes cluster

a. masters
b. minions

46. What is the role of master ?

Ans: Master node is a single node that manages the entire cluster.It coordinates activities like scheduling applications , maintaining applications , scaling applications and so on.

47. What is the role of minions ?

Ans: Minions are the virtual machines that runs containers having maching learning service. They serve as worker machines.

48. In minion node who manages the node and communicates with master ?

Ans: Each minion node contains a kubelet, the agent which manages the node and communicates with master.

49. What is kubernetes deployment ?

Ans: A Kubernetes deployment is all about running a set of pods, which implements a service.The specification of deployment are mentioned in a configuration file generally a .yaml file.

50. How auto scaling of containers can be achieved ?

Ans: Using kubectl autoscale command