**1. Hadoop Installation (Single Node/Multiple Node)**

**Questions:**

1. What are the system requirements for single-node and multi-node Hadoop installation?
2. What configuration files need to be edited in Hadoop?
3. How do you format the Hadoop Distributed File System (HDFS)?
4. What’s the role of NameNode and DataNode in HDFS?
5. How is pseudo-distributed mode different from fully distributed?
6. Which Java version is compatible with your Hadoop version?
7. What are common issues faced during multi-node setup?
8. How do you start and stop Hadoop daemons?
9. How to verify Hadoop installation is successful?
10. What commands are used to put and get data from HDFS?

**2. MapReduce Log File Processing (Java)**

**Questions:**

1. How do you parse a log file in Java?
2. What are the key components of a MapReduce program?
3. How do you define the Mapper and Reducer classes?
4. How do you convert a log entry into a key-value pair?
5. How to calculate total logged-in duration per user?
6. What is the use of context.write() in MapReduce?
7. How do you run a MapReduce job in pseudo-distributed mode?
8. How do you handle invalid log entries?
9. What kind of sorting/comparator logic can be used?
10. How do you view the output of a MapReduce job?

**3. HiveQL Flight Info System**

**Questions:**

1. How do you create a Hive database and table?
2. What is an external table in Hive?
3. How do you load a CSV file into Hive?
4. How do you alter table schemas in Hive?
5. What is the syntax for inserting new values?
6. How do you join multiple tables in Hive?
7. How do you create and use indexes in Hive?
8. What are partitions and buckets in Hive?
9. How to calculate average delay using HiveQL?
10. How do you query data efficiently using Hive?

**4. Python Operations on Facebook Metrics Dataset**

**Questions:**

1. How to load Facebook metrics data using pandas?
2. How do you create subsets using filter conditions?
3. How do you merge multiple dataframes?
4. What are the steps to sort a DataFrame?
5. How do you transpose a DataFrame?
6. How do you reshape data using melt/pivot?
7. How do you handle missing values?
8. What is the use of groupby() in summarizing?
9. How to calculate correlation among features?
10. How to export the final processed data?

**5. Python on Air Quality and Heart Disease Data**

**Questions:**

1. How do you clean missing and invalid values?
2. How do you integrate two different datasets?
3. How do you apply transformations like scaling?
4. How do you detect and correct outliers or errors?
5. How do you split data into train/test sets?
6. What models are suitable for heart disease prediction?
7. How to evaluate model accuracy?
8. How to handle imbalanced data?
9. How to encode categorical features?
10. How to save and load models using joblib/pickle?

**6. Python Visualization (Matplotlib, Seaborn)**

**Questions:**

1. How to visualize air quality over time?
2. What seaborn plot is suitable for correlation heatmaps?
3. How do you use subplots to show multiple variables?
4. How to plot boxplots for comparing groups?
5. What is the best way to show trends using line graphs?
6. How do you add titles and labels to plots?
7. How to annotate values on plots?
8. How to customize figure aesthetics?
9. How to combine multiple plots into one figure?
10. How to save figures as PNG/PDF?

**7. Tableau Visualizations on Adult & Iris Dataset**

**Questions:**

1. What is 1D visualization and how do you do it in Tableau?
2. How to create 2D scatter plots in Tableau?
3. Can Tableau represent 3D data? How is it visualized?
4. How to create a time-series (temporal) chart?
5. How to handle multiple dimensions in a single view?
6. How to build hierarchical (tree) diagrams in Tableau?
7. How to visualize network data in Tableau?
8. How to filter and drill down data interactively?
9. How do you blend data from different sources?
10. How do you publish and share Tableau dashboards?

**8. Python E-commerce Review Scraper**

**Questions:**

1. Which site are you targeting for scraping?
2. How to inspect elements to find review data?
3. How to handle dynamic loading (AJAX/JS) using Selenium?
4. What libraries are used for scraping (BeautifulSoup, requests)?
5. How to avoid getting blocked (headers, sleep, proxies)?
6. What fields are you extracting (comment, name, rating)?
7. How do you store data (CSV, JSON, DB)?
8. How to scrape multiple pages automatically?
9. How to detect and remove duplicates?
10. How to analyze and visualize scraped reviews?