1. How would you handle streaming. Give suggestions about architecture, Technologies, and pipelines that what you would use as Solutions. You can use architecture diagrams and sketches if you like.

2. Discuss a optimal strategy of how to handle big data from streaming (track events from platform usage). Give examples of architecture, Technologies, and pipelines that you would use.

3. Give an example of how you would automate a process using serverless Jobs likes AWS lambdas. Discuss the specific example of how would you handle tables in S3 using table partitions to optimize storage and to be cost effective.

4. Give a definition and an example of a supervised learning algorithm. Discuss in detail, give a simple example and show some simple code. Explain how would you create a pipeline for this model, and what Technologies would you use. How would you orquestrate the flow, and what type of job you would use to run the machine learning algorithm as cost effective as you can.

5. Give a definition and an example of a unsupervised learning algorithm. Discuss in detail, give a simple example and show some simple code. Explain how would you create a pipeline for this model, and what Technologies would you use. How would you orquestrate the flow, and what type of job you would use to run the machine learning algorithm as cost effective as you can.

6. Discuss everything you know about airflow. Give examples of how you used this technology in real projects from Your previous experience as a data engineer.

7. Give na example of an ETL using mockup or real data. Use at least one join, one Where clause, one group by, and at least three special functions such as rank, max, mean, sum, avg, so on and so forth. You can use Pysql, or pandas if you will. Explain how you would create a gold datalake to a Analytics layer for example a tableau server. Use Your github to showcase Your answer, and use some API to store the data (for example Google API, or even a google drive personal account)

8. Explain the subtle differences between ETL and ELT. Give theoretical examples, and real examples from Your previous experience as a data engineer. Give details of the Technologies, and the pipeline that you used or helped create on the Project you are describing.

9. Explain what a schema is. Explain the difference between datawarehouses, data lakes, and serverless Jobs and servers. Discuss how would you handle big data, and which Technologies would you use. Discuss cases using spark, scala, and SQL. What are Your views on current big data Technologies?

10. Given the two tables A and B, give the result for the query below

Table A

ID	Name	Born_in
9991	John	Liverpool
9992	Paul	Liverpool
9993	Ringo	Liverpool
9994	George	Liverpool

Table B

ID	Instrument	Football_Club
9991	Vocal	Liverpool
9992	Base	Liverpool
9993	Drums	Liverpool
9995	Guitar	Arsenal
9992	Piano	Liverpool

```
SELECT A.ID, A.Name, A.Born_in, B.Instrument, B.Football_Club
From A
LEFT JOIN B
ON A.ID = B.ID;
```