Databricks Questions

- **Ramakrishna Gunimanikala**

1) What is Databricks

2) Different types of clusters in Databricks

3) What are cluster pools

4) When do you use job cluster/Interactive cluster

5) Cost comparison between interactive/job cluster

6) Different ways to submit spark job to Databricks cluster

7) How do you pass parameters to Databricks notebook

8) How do you install libraries in Databricks cluster

9) What are init scripts in Databricks

10) How to get run id and job id of a Databricks notebook

11) Difference between dbutils.notebook.run / % run

12) Use cases of dbutils.notenook.run vs % run

13) What is DBFS ?

14) Different ways to access ADLS gen2 from Azure Databricks

15) How do you mount ADLS container to DBFS

16) Databricks cost optimization techniques

17) How do you trigger one notebook from another notebook ?

18) How to return output of Databricks notebook

19) How do you access secrets from Databricks

20) What is auto scaling n which algorithm does Databricks uses to upscale/downscale workers

21) Which cluster manager does databricks uses for providing resources to spark jobs

22) How many executors does databricks cluster launches by default

23) How do you overwrite default number of executors in Databricks

24) Difference between delta cache and spark cache

25) Difference between high concurrency cluster and standard cluster

26) Types of Databricks workspaces

27) How do you import notebook to workspace

28) How do you schedule Databricks job

29) Does Databricks supports event trigger

30) Default expire time of Databricks token

31) How do you generate life time validity token

32) How to run two command cells parallelly in Databricks notebook

33) Which meta store does Databricks use

34) How to configure customer meta store in Databricks

35) How do you deploy notebooks to Databricks Workspace using CICD process

36) How to configure Databricks cli

37) How do you list files using Databricks cli

38) How do you list/delete files from Databricks UI

39) Difference between Azure Databricks n Azure Hdinsights n which one do you prefer

40) What is Delta lake

41) Explain how Databricks able to achieve ACID properties

42) What is data skipping

43) What is optimize n use cases

44) What is Vacuum n use cases

45) What is Z-ordering

46) SCD1 using Databricks merge command

47) SCD2 using Databricks merge command

48) What is Delta table

49) What are operation metrics in delta lake

50) How do you check how many versions of data your delta table has

51) How to revert the data to previous version

52) How do you delete last 6 versions data

53) By default how many versions does delta table stores

54) What is _delta_log directory n what info does it contains