



Mission100 Azure Data Engineer Course By Deepak Goyal

<https://adeus.azurelib.com>

Email at: admin@azurelib.com

Ask Queries here: <https://www.linkedin.com/in/deepak-goyal-93805a17/>

classmate
Date _____
Page _____

Wipro 1st Round:

1. What is pickle function in python.
2. Write a code to fetch 1, 3, 5, 9 rows from dataframe.
3. How do you change the file details at Storage level.
4. How do you implement security in cloud.
5. How do you fetch files from ADLS.

ADLS has 5 folders, each folder contains 4 files each, you need to fetch last modified file from ADLS and move rest of the files to Archive folder.

Write code to join 2 tables

Match	Country
1	India
2	Aus
3	NZ
4	SA

Write a SQL query to prepare a Data such that each team should play one match with other team, but there should not be vice-versa.



Mission100 Azure Data Engineer Course By Deepak Goyal

<https://adeus.azurelib.com>

Email at: admin@azurelib.com

Ask Queries here: <https://www.linkedin.com/in/deepak-goyal-93805a17/>

HCL 1st Round:

- ① SQL query to check occurrence of 'i' in the string "INDIA"
- ② how do you implement incremental data load
- ③ Can you do Nested for each loop in ADP.
- ④ what transformation you have used in Notebook
- ⑤ which all triggers you have used
- ⑥ how do you handle your deployments!
- ⑦ how can you performance tune your copy activity
- ⑧ how do you create a trigger which runs on Month last day.
- ⑨ Suppose you have FTP, once the file dropped to FTP you need to run the pipeline to load data
- ⑩ how do you Analyse any requirement and implement (Step to Build pipeline, Point form Requirement Analysis)
- ⑪ performance optimization in Databricks
- ⑫ which lang you are comfortable to code in Databricks
- ⑬ what is SCD, explain types

Course Link: <https://adeus.azurelib.com>

Email at: admin@azurelib.com

Azurelib Academy By Deepak Goyal

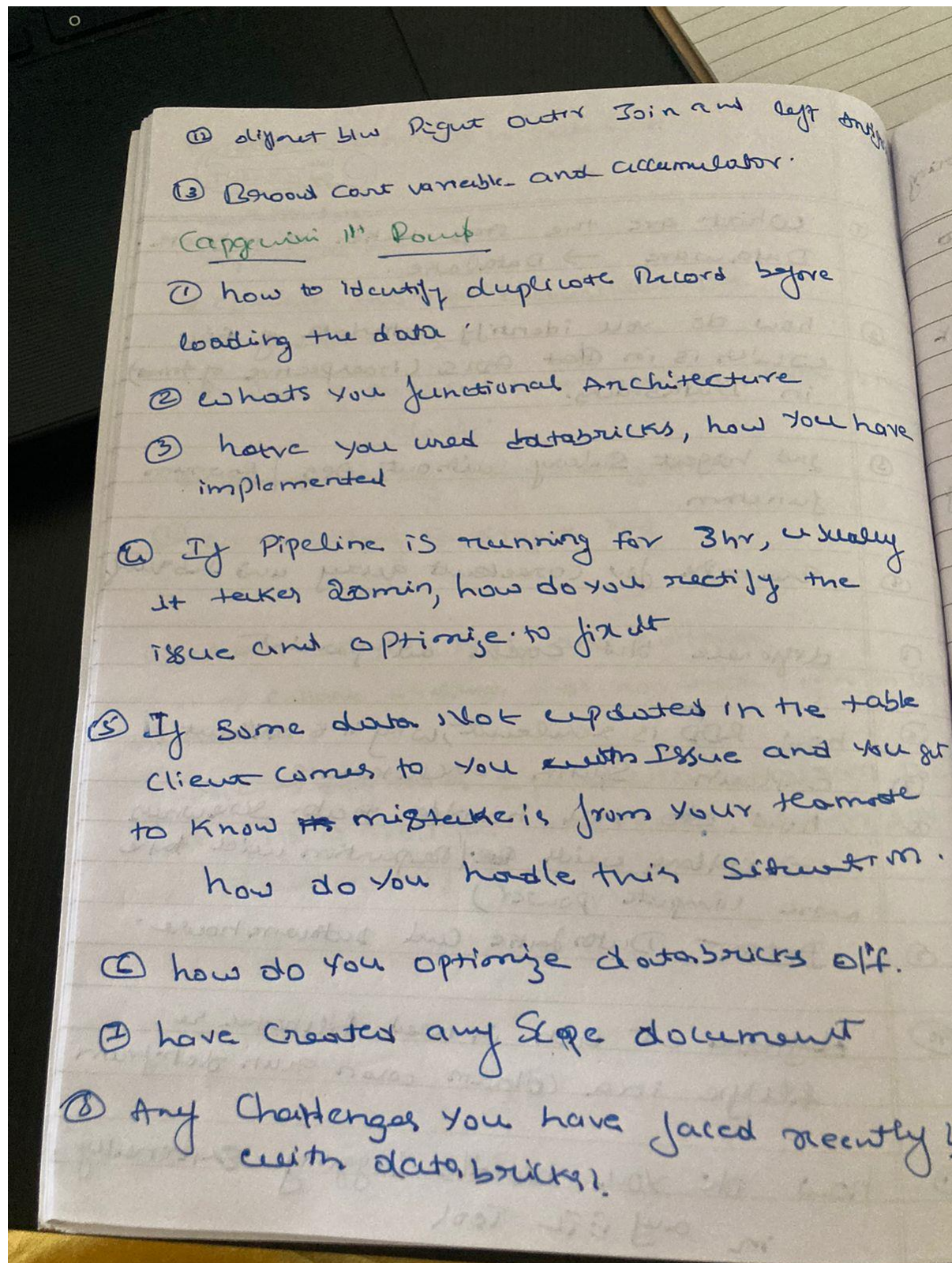


Mission100 Azure Data Engineer Course By Deepak Goyal

<https://adeus.azurelib.com>

Email at: admin@azurelib.com

Ask Queries here: <https://www.linkedin.com/in/deepak-goyal-93805a17/>



Course Link: <https://adeus.azurelib.com>

Email at: admin@azurelib.com

Azurelib Academy By Deepak Goyal

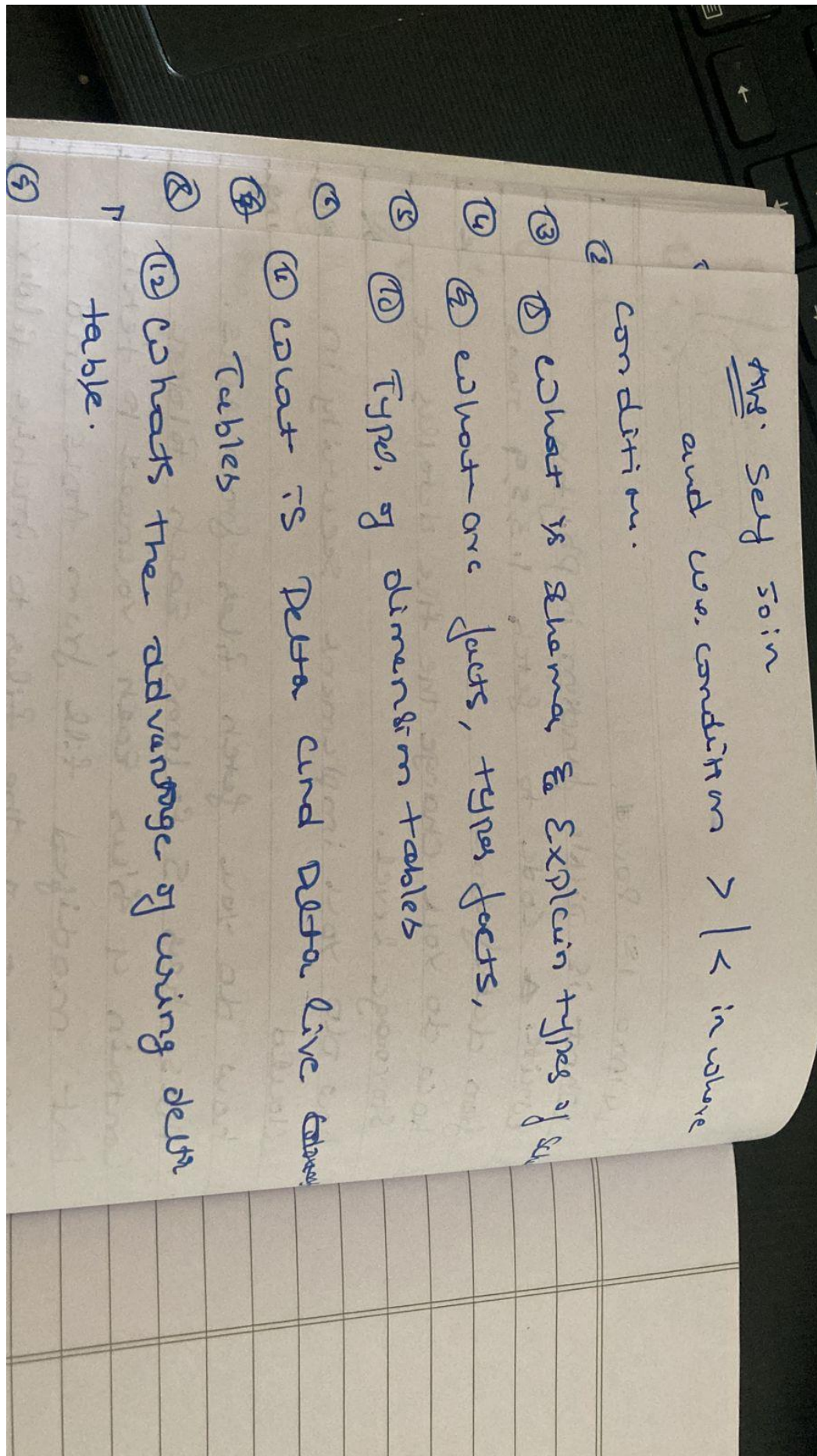


Mission100 Azure Data Engineer Course By Deepak Goyal

<https://adeus.azurelib.com>

Email at: admin@azurelib.com

Ask Queries here: <https://www.linkedin.com/in/deepak-goyal-93805a17/>





Mission100 Azure Data Engineer Course By Deepak Goyal

<https://adeus.azurelib.com>

Email at: admin@azurelib.com

Ask Queries here: <https://www.linkedin.com/in/deepak-goyal-93805a17/>

```
Untitled - Notepad
File Edit Format View Help
Q1 - - ADF SCENARIO BASED Q :
There is a Pipeline which loads Data from On Prem Source to Target - Approx 1 L Records Daily
Requirement : From tomorrow Onwards Clients wants You to LOAD ONLY THOSE Records which are NEW in the Source System .
Eg - Out of 1L Records , Only 100 Records have Changed as NEW & CHANGED Records
Build a Pipeline Solution so that You ONLY LOAD THOSE 100 NEW Records from Source System .
Note - There is NO DATE /TIMESTAMP Column

Q2: How do you Implement SCD -3 based Data Load for a CSV Type Source Data which has NO TIME DATE Column ?

Q3 - How do you Implement CDC Based data Loading for a STRUCTURED Database using ADF ?

Q4- : : : DATABRICKS / PYTHON / PYSPARK - CODE TEST

Replace the First 2 Occurrence of word BLR with KOL from Below Dataset

1 SS 12 1234 10/12/2022 BLR
2 SS 12 1234 10/12/2022 HYD
3 SS 12 1234 10/12/2022 DEL
4 SS 12 1234 10/12/2022 BLR
5 SS 12 1234 10/12/2022 BLR
6 SS 12 1234 10/12/2022 BLR
7 SS 12 1234 10/12/2022 HYDF
8 SS 12 1234 10/12/2022 PUN

NOTE - This is CSV File & Last Column = CITY

Q5) Explain DAG Architecture and it's impact to repartition & coalesce?

Q6) How do you call multiple notebooks from 1 activity in ADF?

Q7) Write a query to fetch nth occurrence of a number in repeated data set?
E.g : 1,3,4,4,4,2,4,2,6,8,4,3,3

Q8) How to source data from sftp/Snowflake connectors using descriptors?
```

Course Link: <https://adeus.azurelib.com>

Email at: admin@azurelib.com

Azurelib Academy By Deepak Goyal

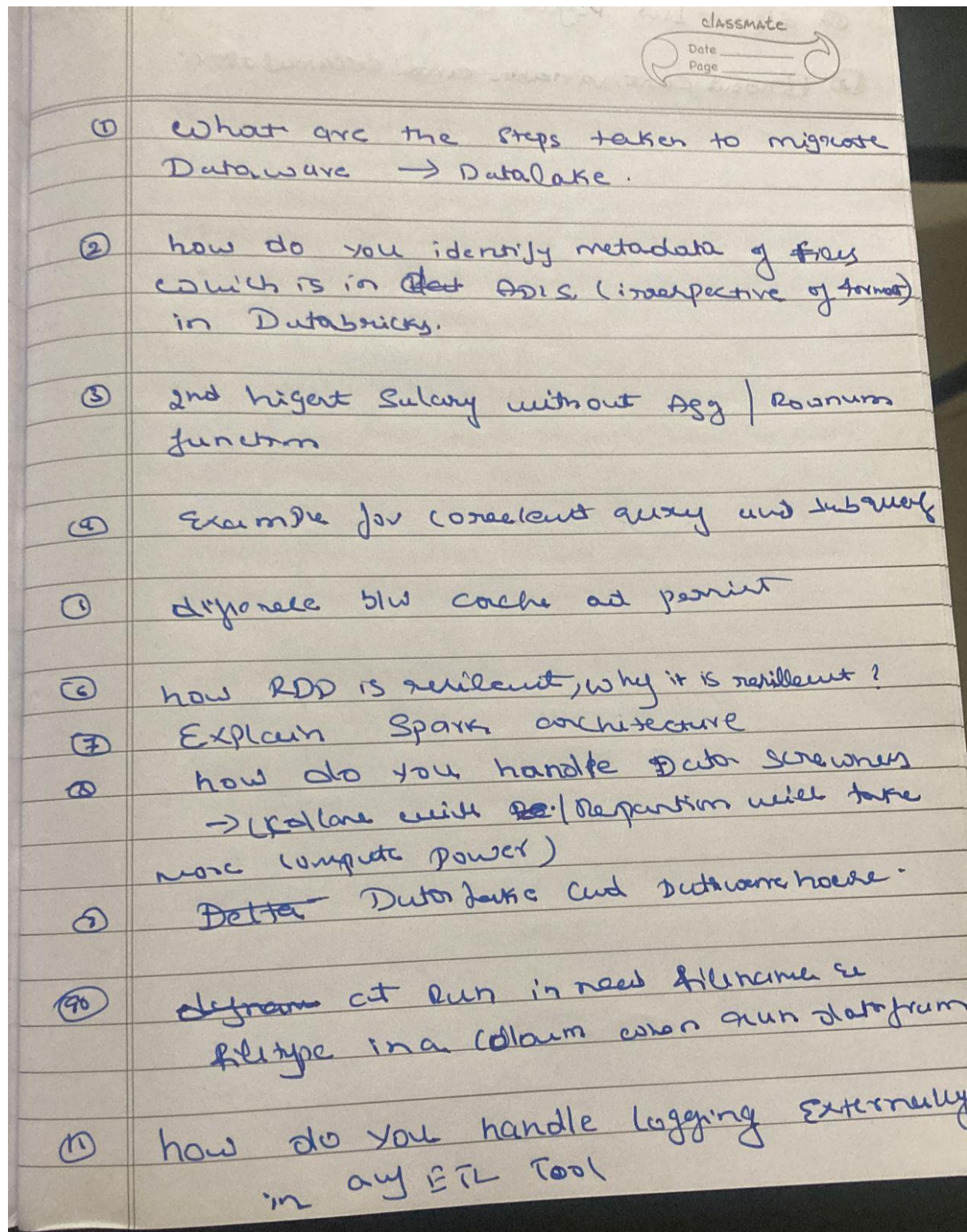


Mission100 Azure Data Engineer Course By Deepak Goyal

<https://adeus.azurelib.com>

Email at: admin@azurelib.com

Ask Queries here: <https://www.linkedin.com/in/deepak-goyal-93805a17/>



Course Link: <https://adeus.azurelib.com>

Email at: admin@azurelib.com

Azurelib Academy By Deepak Goyal



Mission100 Azure Data Engineer Course By Deepak Goyal

<https://adeus.azurelib.com>

Email at: admin@azurelib.com

Ask Queries here: <https://www.linkedin.com/in/deepak-goyal-93805a17/>

KPMG interview questions
Interview Questions

Q1. Difference between RDD, Dataframe and Dataset. How and what you have used in you databricks for data anlysis

Q2. What are key components in ADF? What all you have used in your pipeline?

Q3. Do you create any encryprion key in Databricks? Cluster size in Databricks.

Q4. Difference between ADLS gen 1 and gen 2?

Q5. What is Semantic layer?

16:24



Mediolanum

- 1.difference between partition and group by
2. Difference between blob storage and data lake
3. How you implement security in cloud
4. How you clone / replicate data from on premise to cloud (migration)
5. 2nd highest amount from order table
6. What is difference between data mart and data cube
7. Normalisation
8. How design dynamic pipeline and handle /run dynamic pipeline in adf
9. Have you used API
10. How do you manage deployments
11. Do you have experience in monitoring
12. How do you optimise pipeline
13. What is composite key
14. Have you worked on incremental load
15. How do you handle data loading
16. How you do handle sql injections
17. How do you handle security on cloud data

19:40



Course Link: <https://adeus.azurelib.com>

Email at: admin@azurelib.com

Azurelib Academy By Deepak Goyal



Mission100 Azure Data Engineer Course By Deepak Goyal

<https://adeus.azurelib.com>

Email at: admin@azurelib.com

Ask Queries here: <https://www.linkedin.com/in/deepak-goyal-93805a17/>

Mastercard 2nd round

Difference b/w hive and spark?

Broadcast join?

How to fetch data from hdfs?

2 sql queries?

Hive query optimisation?

How you will test your data?

There were other questions but do not remember them. 17:37



Capgemini questions - most of the question related to project. 2 functions used in ADB in your project 3. Optimization used in adb 4. How to optimise api which we calling in ADF ,if it failed due to huge load 5. How to implement scd type 3 in adb 6. Left semi join. 15:41

Course Link: <https://adeus.azurelib.com>

Email at: admin@azurelib.com

Azurelib Academy By Deepak Goyal



Mission100 Azure Data Engineer Course By Deepak Goyal

<https://adeus.azurelib.com>

Email at: admin@azurelib.com

Ask Queries here: <https://www.linkedin.com/in/deepak-goyal-93805a17/>

Kpmg interview questions

How will you read data with multiple gb from sftp in adf?
CICD?
Schema inforcement in delta table?
Trigger on sftp server using adf when data arrives?
What happens if driver machine fails in databricks?
Memory management in spark?
Simple sql queries ?
View vs materialized view?
Etl pipeline architecture?
Spark architecture?
Different types of join in spark?

19:16

👍 7

Mastercard interview questions asked

Bucket and partitioning?
SCD Type?
How old data you store in data warehouse if we talk about scd 2 ?
Data mart?
Hdfs?
Testing framework?
4th highest salary?
View vs materialised view?
Union and union all?
Did you implement any data warehouse?
Can we have partitioning and bucketing together?

19:13

👍 8

Course Link: <https://adeus.azurelib.com>

Email at: admin@azurelib.com

Azurelib Academy By Deepak Goyal



Mission100 Azure Data Engineer Course By Deepak Goyal

<https://adeus.azurelib.com>

Email at: admin@azurelib.com

Ask Queries here: <https://www.linkedin.com/in/deepak-goyal-93805a17/>

Questions asked in Cognizant interview

Databricks runtime used?

MAP join?

Search a text in big file?

How you will setup a etl pipeline?

Fetch 2 csv files , remove duplicates and then join?

What are magic commands?

How to run job in databricks?

How to count number of executors?

Challenges face in your projects?

Repartition vs coalesce ?

Spark sql?

Second highest salary?

What is equi join?

sparkSession vs sparkContext?

Tuple, dictionary?

12:08



Course Link: <https://adeus.azurelib.com>

Email at: admin@azurelib.com

Azurelib Academy By Deepak Goyal



Mission100 Azure Data Engineer Course By Deepak Goyal

<https://adeus.azurelib.com>

Email at: admin@azurelib.com

Ask Queries here: <https://www.linkedin.com/in/deepak-goyal-93805a17/>

Fractal L1

1. SQL question to print max salary department wise along with Emp Salary
2. Implement same solution using window functions
3. Python program to reverse an integer
4. Python program to print if a number is prime
5. What is Databricks
6. What is a driver and executor node
7. Difference between blob and ADLS
8. Why do we need Databricks
9. What are the types of clusters
10. Can we iterate over a dataframe using loops
11. What is integration run time and it's types
12. How do we get to know if a pipeline has failed or succeeded in ADF
13. What is a Delta table
14. What are fact and dimension tables
15. What are the different ways of reading a file from databricks present in ADLS
16. Can we run a Pandas code in databricks and if yes which is a better option...
Running the same dataset on Pandas or PySpark.

14:19

👍 13

Course Link: <https://adeus.azurelib.com>

Email at: admin@azurelib.com

Azurelib Academy By Deepak Goyal



Mission100 Azure Data Engineer Course By Deepak Goyal

<https://adeus.azurelib.com>

Email at: admin@azurelib.com

Ask Queries here: <https://www.linkedin.com/in/deepak-goyal-93805a17/>

Capgemini- Interview - May27th

=====

- 1) Could you provide a brief introduction about yourself and your project as an Azure Data Engineer?
- 2) How can PySpark be utilized to establish a connection with Azure Data Lake Storage (ADLS), retrieve data from a CSV file, convert it to Parquet format, and store it back in ADLS?
- 3) How can the same functionality be implemented simultaneously in two separate notebooks within Databricks?
- 4) Can you elaborate on the two available types of clusters in Databricks?
- 5) What are the different types of Delta Lake tables?
- 6) Can you explain the roles and functions of each layer in the three-layer architecture of Databricks?
- 7) Within Databricks, in which layer should production data be ingested?
- 8) When retrieving records from Table B while including the common records between Table A and Table B, what type of join should be used?
- 9) What are the differences between an Inner Join and a Self Join? Could you provide an example to illustrate each?
- 10) What is a Cartesian join, and how does it differ from a self-join and a full outer join? Additionally, could you provide an example to further explain, especially when null records are present in one table?
- 11) Can you provide an explanation of window functions in SQL?
- 12) How do the lead and lag functions in SQL differ from each other?
- 13) What is the concept of normalization, and what are the different normalization types?

Course Link: <https://adeus.azurelib.com>

Email at: admin@azurelib.com

Azurelib Academy By Deepak Goyal



Mission100 Azure Data Engineer Course By Deepak Goyal

<https://adeus.azurelib.com>

Email at: admin@azurelib.com

Ask Queries here: <https://www.linkedin.com/in/deepak-goyal-93805a17/>

Hexaware: May 22-2023

=====

- 1)Can you provide a brief overview of your background as an Azure Data Engineer?
- 2)What are your specific responsibilities and roles as an Azure Data Engineer?
- 3)Could you explain the reasons for using Databricks and its practical applications?
- 4)Can you highlight the key distinctions between Azure Data Factory (ADF) and Databricks?
- 5)In a scenario where I need to move 100 tables from a source SQL database to a target SQL database, and the number of columns in the source and target tables differ, what approach should I take to accomplish this task using Azure Data Factory?
- 6)Can you elaborate on the differences between blob storage and Azure Data Lake Storage (ADLS)?
- 7)What does the term "managed table" refer to in the context of PySpark?
- 8)Could you outline the disparities between repartitioning and coalescing in Spark?

21:28



Course Link: <https://adeus.azurelib.com>

Email at: admin@azurelib.com

Azurelib Academy By Deepak Goyal



Mission100 Azure Data Engineer Course By Deepak Goyal

<https://adeus.azurelib.com>

Email at: admin@azurelib.com

Ask Queries here: <https://www.linkedin.com/in/deepak-goyal-93805a17/>

Mediolanum

1. difference between partition and group by
2. Difference between blob storage and data lake
3. How you implement security in cloud
4. How you clone / replicate data from on premise to cloud (migration)
5. 2nd highest amount from order table
6. What is difference between data mart and data cube
7. Normalisation
8. How design dynamic pipeline and handle /run dynamic pipeline in adf
9. Have you used API
10. How do you manage deployments
11. Do you have experience in monitoring
12. How do you optimise pipeline
13. What is composite key
14. Have you worked on incremental load
15. How do you handle data loading
16. How you do handle sql injections
17. How do you handle security on cloud data

11:14



Course Link: <https://adeus.azurelib.com>

Email at: admin@azurelib.com

Azurelib Academy By Deepak Goyal