Accenture

1. What optimisation techniques used in BQ (on gold layer)?
2. Types of Joins?
3. Diff b/w Left Join & Right Join?
4. a) Table:

   Emp id, Emp name , dept id

   b) Dept:

   Dept id, dept name

   Write a left join query?

5. What is CTE?
6. Third Highest salary query using CTE?
7. Diff b/w Rank & Dense Rank?
8. Store Procedure in SQL?
9. Diff b/w store procedure & view?
10. Use of having in SQL?
11. Surrogate key, Primary key vs Surrogate key?

Altimetrik

1. Self-introduction?
2. What type of work you are doing in GCP?
3. What type of files you are receiving?
4. Storage classes you used in project.
5. What is the largest file size you handle?
6. When we run the dag what type of error?
7. Reverse the string in python without inbuilt functions or higher order function.
   Str_var="abcde"
8. Second max of the salary?
9. Replace the tablex column from col1 to col2?

Altimetrik

1. Tell me about yourself and Explain previous Project?
2. In GCP what are the services your used?
3. How you rate in BigQuery?
4. Orc file to Bigquery loading?
5. Do you know BQ commands?
6. How do you write query using CLI?
7. How do you get DDL of the bigquery table?
8. To get row count of all in BQ?
9. I need to see DDL of each table in BQ Console?
10. Foe each partition how many records are there?
11. How to create partition id & Partition Summary?

Cloud Data Engineer – GCP
Instructor – Srinivas
Interview Questions – Company wise

CELEBRATING
14 YEARS

QualityThoug

12. What are the column based we create based?
13. How to Authentiate GCS?
14. How you get Authenticate in GCS?
15. Where service Account will store in GCP?
16. GCP to Local how do you we copy file?
17. How to edit file in GCS?
18. What are the services your used in GCP?
19. What is DAG and what are the operator used?

Altimetrik

1. Self-Introduction?
2. Source of data?
3. Partition we use before using BigQuery?
4. How will use reduced the cost in the bigquery in clustering and partition where we use (stage and history)?
5. Customer& Syntax?
6. I have a table USA,UAE have the get the data for 2021 Jan to Dec 2022 customer?
7. Python knowledge
   Tuple= (1,2,4,5)
   U can insert 3 after 2?
8. Different between list and tuple?

Altrimax

1. We need userid,course-enrolled,enrolment.TS,cort-USD

Users-stats
Course-enrolled
Order-id
Entrollment-TS
Cort-USD

Users
User-id
User-name
Registration - Ts
emailid

We need user_id,   user_name,course_entrolled enrolment_Ts, Cost_USd

2. Need userid,amount_USD spending more than 100 USD?
3. What is DAG,explain?
4. How good at python in DAG execution?
5. How comfortable with dashboard you used?
6. Can you create dashboard?

Scanned with OKEN Scanner

Cloud Data Engineer – GCP
Instructor – Srinivas
Interview Questions – Company wise

CELEBRATING
14 YEARS

QualityThought

Altimetric

1. what are the services you have used in gcp?
2. what is the difference between partition and clustering?
3. how does increase the performance?
4. how does sorting helps in query performance in big query?
5. how does arranging clustering helping the query performance?
6. how many clusters column can i have in a table.
7. how many partitioned columns can i have?
8. how many numbers of partitions can i have?
9. if i load daily basis you get a table with 20 years of data how can i load it with partions
10. i have 1000 table in dataset i need row count of each table how can i get it?
11. how are you creating table in bq?
12. What type of table will create?
13. What is the differnce between internal table and external table?
14. What are bq commands a dataengineer use in his daily work?
15. How are you authenticating to gcs bucket?
16. If you want a place in gcs bucket, how can you place it from where to where?
17. If i want to upload multiple files at a time in gcs buckets in parallelly how can i do?
18. If i want to edit content of the file how can i do?
19. How do you load data from gcs to bigquery?
20. How are you running from gcs to bq?
21. Are you creating dags? how do you create dags?
22. Do you have experience in pyspark?
23. Do you have experience in python?
24. What kind of your things have done in python in gcp?
25. As i am getting orc file how can i delate first row and last row in python?

Altritmatrix

1. Migration process in the project?
2. What connection is used from on previous to GCP?
3. Experience in GCP?
4. What are compoments in GCP you are worked?
5. Any schedule job you done?
6. BQ ,what is the use of partitioning?
7. I want optinization in BQ for two column dat column and country column. Is it possible?
8. stacs in Airfow, explain each star in Airflow?
9. Cost component how to findout and view details?
10. On-Premises which component you have knowledge?
11. UI on G-cloud command which method you using in console?
12. How much volumes of data your working?
13. What kind of data your using in the project?
14. Have you worked on dashboard ,what kind of scripting method your using?
15. Google visualization tools used in the project?
16. Aware of SQL quries?
17. What is the use of windowing function?
18. Difference between Rank and Dense Rank?

19. How you can instant the packages in Python?
20. What are the packages in the python?
21. Security tools on packages in the python?
22. Quality checking in the code?
23. CTE command in SQL?
24. What is the use of CTE?
25. Knowledge in UNIX Command?
26. Copy data from one location to another?
27. To search a text or word in a file.which UNIX command is used?
28. What the use of set Command?

Annalect

1. Tell me about yourself and Explain your Previous Project?
2. Name of Employees whose age >30?
3. What is the difference between Primary Key and Foreign Key?
4. Table A has (PK) Primary Key and Table B Foreign Key in another table so I want to delete table A?

Apexon

1. Input='aaabbbaabbc'
   Output='3a3b2a2b2c'} Python and pyspark
2. Rank & Dense Rank?
3. How to read a CSV file using pyspark?
4. Tell me the BiqQuery Architecture?
5. What is dremel?

Clairvoyant

1. Brun down & Brun up chat in Agile?
2. Sos in Agile?
3. Sprint backlog & product backlog in Agile?
4. Row ID vs Row Num?
5. Delete vs Trucate?
6. function vs Procedure in SQL?
7. Where is Having in SQL?
8. What is Surrogate key?
9. What is velocity in Agile?
10. Rank vs Dense Rank vs Row number in SQL?

Clairvoyant

1. Experience in python?
2. What are oops functions in python?

3. What is Pep 8 ?
4. What are decorators?
5. What getter and setters?
6. What is range in function?
7. What is lambda function?
8. What is polymorphism in Python (oops)?
9. for I range (1,10,-2):
   Print (i)
   Output?
10. Experience in airflow?
11. Which operators used in airflow?
12. Types of executors airflow?
13. What is use of celery executors?
14. Some workers for specific task, another task. How do you execute?
15. Have you used multiple workers in airflow?
16. How do u define a Dag in an airflow?
17. Moving data of on table to another table. How can you do it from BQ to onpremise?
18. How much datasize you are working on daily basis?
19. How do used file using Spark?
20. How can you wrote a have in spec?

Cognizant

1. Tell me about your Experience in GCP?
2. What is Dataflow and usage?
3. Tell me about Composer?
4. Found Timeout error in Daywhat is Your first approach?
5. How Many versions are there in cloud Composer?
6. Difference b/w Composer 1 and Composer 2?
7. Tell me something about Big Query?
8. Which language is used in Big Query?
9. What are types of joins in SQL?
10. What command you used if you to remove duplicates values?

Cognizant

1. Explain your Recent project?
2. What are the operators used in Airflow?
3. Can we call one dag to another dag?
4. What are the ETL tool you worked?
5. What are the RDBMS worked on?
6. Why would you choose snowflake?
7. What is the role in snowflake?
8. Explain SDLC?
9.   T1   T2

       ID 1   ID 3

Cloud Data Engineer – GCP
Instructor – Srinivas
Interview Questions – Company wise

CELEBRATING
14 YEARS

QualityThough

2	1
3	2
4	5
Null	Null
5	Null
2	1

Select (name, Sal, lead(Sal)) from emp?

10. Difference between OLTP & OLAP?
11. Difference between View & Materialized view?
12. Python Input Keerthy? Output Reverse of string?
13. Have you worked on optimization technique in your project?
14. Why should we choose Airflow instead of data pipeline?
15. What is multireading in python?
16. Diff data frame & dynamic frame?
17. What is the process in snowflake?
18. What is the diff between List & Tuple?
19. What are the storage classes?

Cognizant

1. What was the GCP service you worked upon?
2. What is dataset as Big query?
3. What are different types of Tables?
4. ename,esql.detro, extract ename who is during the high salary?
5. I/P:

Col 1	Col 2
1	a
1	b
1	c
2	f
2	g
3	j

O/P:

Col 1	Col2
1	(a,b,c)
2	(f,g)
3	j

6. What is command which we use to load file to the bucket?
7. Types of Storage Classes?
8. What about Casting in Storage?
9. I want to upload a file to bucket after 10 days I want delete the file automatically.

QualityThought[n]

Deloitte

1. What we do in Big Query?
2. Daily what type of data are you collected?
3. What type of transformations, execution query also explain?
4. Explain nest and unnest?
5. Did you use this functions nest and unnest in your project?
6. Where you use Nest and Unnest?
7. What the role of project?
8. Constraints are used in big query?
9. Per day 7 GB is allowed or not in big query?
10. Do you validate data before moving into big query?
11. What if the data is jumbled? Do you move it into big query without any changes?
12. Single file or multiple? If multiple files how many do you receive?
13. Single file from all sources? Or different files for different sources?
14. Number of Sources?
15. How do you move data from GCS to BQ?
16. What is your client expectation?
17. How many files u receive?
18. Types of portioning?
19. Dimensions in your data?
20. Hierarchy in your data?
21. How do you remove duplicates?
22. Operators used in Dags?
23. How do you update the changes made in source?
24. Have you worked on cloud SQL, cloud spanner, cloud function€/
25. GCS to Big query migration process?
26. Difference between rank and dense rank?
27. How will you running DAG in Airflow?
28. Big query architecture?
29. How will you do partitioning in a table?
30. Partitioning syntax?
31. How to load the data from GCS to Big query?
32. Federated query?
33. Cloud sql and cloud spanner?
34. Write a dag to lad the data GCS to Big Query?
35. Write query getting previous record?
36. Write a python code for febinocci series?
37. Do you know about dataflow and data proc?
38. Difference between normal cluster and data proc cluster?
39. Have u used git hub?
40. What is git stash?
41. What is snapshot in bigquery?
42. How to move file from one path to other path?
43. Have u worked on Jenkins?
44. Minus operators (SQL, Bigquery)?
45. How u check dag errors in airflow?
46. Before moving to deployment, how u approached quality assurance team?

Cloud Data Engineer - GCP
Instructor - Srinivas
Interview Questions - Company wise

QualityThoug

CELEBRATING
14 YEARS

Deloitte

1. Remove all the duplicate records from table?
2. Surrogate Key vs Unique Key vs Primary Key?
3. Materialised views?
4. What ETL tool you have used?
5. DDL, DML, DCL?
6. SCD type 2?
7. How do you load data into BQ?

Deloitte

1. Self?
2. What Storage Classes you used?
3. What Data Retention Policy you are following?
4. If Data or Table in BigQuery get deleted, How to retrieve it?
5. If cost results in Big Query belongs to all users or specific users?
6. If two people using same service accounts, whom can see results?
7. Did you Scheduled the query and is their limitations in that?
8. I have table with 1000 records,i am using limit to 10 records ,if it scans all the records or not?
9. According to cost which table is better CDE or
10. What are temporary table in Big Query?
11. Did you moved from BQ to GCS (bq extract)?
12. What is P-Collections in Dataflow?
13. Did you worked on dataproc,pub/sub?
14. In Composer if you written any dag?
15. Once you fire a query,how it works, in BigQuery?
16. In fixed pricing how many slots available?
17. What are premtable views?
18. Have a table "persontable"

Id	email
1	john@example.com
2	bob@example.com
3	john@example.com

Here,john is repeated so eliminate one john output be like

O/P:-

Id	email
1	john@example.com
2	bob@example.com

19. Table Employee

Id	name	salary	ManagerId
1	Joe	70000	3
2	Henry	80000	4

Scanned with OKEN Scanner

...ud Data Engineer – GCP
...structor – Srinivas
Interview Questions – Company wise

QualityThought[R]

CELEBRATING
14 YEARS

| 3 | Sam | 60000 | Null |
| 4 | Max | 90000 | Null |

Display who salary is more than his manager salary

O/P:- Joe

20. R u working on Dataflow & Dataproc?
21. Where you do transformations?
22. Where do you run pyspark (r u working on pyspark or some other team)?
23. What are data frames?
24. Python operator, How to write python script?
25. In cloud Composer?
26. Dependencies in Airflow?
27. Partitioning and significant in Big Query
28. What is Clustering?
29. Differences Between rank and dense rank?

DSquare

1. What Kind of data you worked till now?
2. Which tools did you worked?
3. Difference data scientist and data engineer?
4. What are supervisory validation technique you used in data engineer?
5. Star Schema and snowflake schema?
6. What are ETL tool you worked?
7. Alteryx, Tableau tool in your project?
8. Data request first couple of questions you need to ask before new project start?
9. Key jobs is the data engineer?
10. Difference between GCP and Snowflake?
11. What are advantage of Bigquery and snowllake?
12. Explain map redile?
13. Experience in python scripts, where did you used in real time?

Excelra

1. How to load data from GCS to BigQuery?
2. What files did you worked?
3. What are size of the file?
4. Trigger the BQ load Command?
5. Which operator is used for BQ load?
6. How do you load data is stage and history table?
7. What is External table?
8. How do load file from on premises?
9. GCS with a huge Size?
10. R u families with util Commands?
11. How do you move filefrom on premises to GCS?
12. Where do you execute the command SDK or cloud shell?

Cloud Data Engineer – GCP
Instructor – Srinivas
Interview Questions – Company wise

13. How do you process the file moved from GCS to BQ on Dataflow?
14. Can we save array datatype in BQ?
15. Nested data in BQ?
16. Big Data?
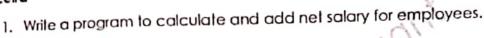17. We have the following tables:

> Employee (empid, empname, Salary, deptid)
> Department (dept-id, dept-name)
> Questions:
> Find Top 5 Salaried employees from each department and fort then in descending order of their salaries, department wise.

18. How do you define Schema for the Phone numbers having multiple numbers?
19. Flattern in emp-name, contact number the function name is used?
20. Which DB did you work?

Excelra

1. Write a program to calculate and add net salary for employees.

Empid	Empname	deptid	salary	bonus percentage
1001	XYZ	10	25000	8.5
1002	abc	20	34000	6.5

2. How do you create the partitions in BQ?
   Integer range partition
   Base partition
   Injection partition
3. How do your secure data for one country from others?
4. Highest salary of (Top-3) each department?
5. What to find no.of alphabet's' in your name?
6. Json?
7. Data Lake vs Data warehouse?

Excelra

1. Self Intro?
2. How did data move from on premises to GCS?
3. How did move the data from GCS to BQ?
4. How did the rev tanks in Airflow?
5. Apart of BQ.Airflow what else other components have worked?
6. What is partition in BQ?
7. Hadoop?
8. Hadoop Architecture?
9. Cross questions about Hadoop & MySQL?
10. What types of databases you have worked?
11. What type of Database in BQ?

12. We have 2 flat files
    EMP.data(having 1 million records)Schema:
    Empid, empname, deptid, Satarlary? Dept.data (having 20 records)Schema:deptid
    deptname.

    Both the flat files do not have Headers and the data is separated"\".

Ford

1. What is Clustering?
2. a1b2c3d4 write a python code print only 'abcd'?
3. {id:1:{a,b,c}} from this json file extract a,b,c?
4. What is Record datatype?
5. Bigquery transformations?

Fossil

1. What is Dataflow?
2. What is Composer?
3. Difference B/W CSV & Parquet?
4. Which is Dag dependencies?
5. From Bucket to Big query explain Airflow code for take file from Bucket to Bigquery?
6. How to you handle null values?
7. CSV-100 words
8. Find the count by using Python Program?
9. Code for reverse string?
10. Partition in Bigquery?

Fossil

1. Rate your SQL & Python?
2. A A B

   | | |

   | | |

   | |

   Inner join

   Left join

   Left outer join

3. Write a query for 5th highest marks of a Table student column.
   Student-id,Student name,m1,m2,m3
4. What are co-related queries?
5. What is CTE into use?
6. What is SDLC and its types?

7. In Python write code to count number of records present in CSV file?
8. Do you know Pandas?
9. In Python which libraries you used?
10. Which GCP components you worked?
11. What is DAG?
12. Who is stored in BigQuery?
13. What are views?
14. What are authorised views?
15. What is major drawback of cloud functions?
16. By using cloud function can we use load data from CGS to BigQuery?
17. What is storage are available in GCP?
18. for every 2 hours you need to schedule through cron
19. Types of operators in airflow?
20. What is dataflow?
21. CSV in GCS to BQ, Automation process?
22. How will trigger the cloud functions?
23. Something fails in the pipelines, flow will you get to notify?
24. What is apache Spack?
25. In pyspark how do you convert 2 dataframe into 1 dataframe?

HCL

1. What are the GCP Services you Worked so far?
2. How you Monitor your Workflow?
3. How you Monitor your Spark Job Submittion?
4. Stages in Spark?
5. Do you know Narrow transformation and wide transformation? (Example)
6. How many partitions can be created in wide transformation?
7. Have you never worked Column level Security?
8. How you display the sensitive data to the Businesspeople in BigQuery?
9. How you fix reconciliations in BigQuery?
10. Write a Python code for without reverse function reverse the strings?

HCL

1. What is time travel?
2. What is struct in BQ, difference between array and struct?
3. What is window functions? What will you write in Over ()?
4. 6.what is row level and column level database?
5. What kind of transformations you did in BQ?
6. How you will give data to. Business stakeholders?
7. What kind of databases you used?
8. How you load data from source to bg?
9. find cummlative sum of below table

date	cutomer	amount
1/6/2022	c1	10

...d Data Engineer – GCP
...structor – Srinivas
Interview Questions – Company wise

CELEBRATING
14 YEARS

QualityThought

1/7/2022	c1	20
1/8/2022	c1	15
1/9/2022	c1	16
1/10/2022	c1	20

expected output.

date	cutomer	amount	cum_sum
1/6/2022	c1	10	10
1/7/2022	c1	20	30
1/8/2022	c1	15	45
1/9/2022	c1	16	61
1/10/2022	c1	20	81

10. What are the optimization techniques you followed in BQ ?

HCL

1. Architecture of Data?
2. Which langague do you use Dataflow?
3. How are you getting data?
4. Any performance issues?
5. What are the Challenges you are facing in datatime?
6. Do you face any issue of cost effective in Bigquery?
7. Are you doing any partitions?
8. Do you know data-time partition limit?
9. If limit is reached what you will do
11. Join 3 tables
    Inner join
    Filtering data
    Aggregations fun)   let me know what performance you will do
12. which table you will consider as a temp table?
13. Which column did you create Index?
14.

Table A	Table B
1	1
1	1
0	0
Null	2
Null	Null

Join Two table by using Equi Join?

15. What will you do to get null columns in above table?
16. Find cumulative sum of below table?

Scanned with OKEN Scanner

Date	customer	amt
1/6/2022	Hitachi	

17. Rate yourself in BQ out of 10?
18. SQL out of 10?
19. Write SQL query for the below Scenario emp whose SQL is greater than avg SQL of these dept?
20. Select * from products.

Product ID	Product Name
100	Nokia
200	Iphone
300	Samsung
400	Motorola

Select * from Sales:

Sales ID	Product ID	Year	Quantity	Price
1	100	2010	25	5000
2	100	2011	16	5000
3	100	2012	8	9000
4	200	2010	10	9000
5	200	2011	15	9000
6	200	2012	20	9000
7	300	2010	20	7000
8	300	2011	18	7000
9	300	2012	20	7000
10	100	2010	10	5000
11	200	2010	50	9000

Top Selling product every year

21. What is Dremal?
22. Colosus

How do connect datatype string to data.

Do you have experience Python?

Do you have experience Tabulea?

Hitachi

1. Can you briefly tell about yourself?
2. Which ETL Tool you used?
3. How to convert one data to another data type?
4. Various ways Big Query can access?
5. What are the packages to install to connect to BQ?
6. In which are the Scenario window functions is used?
7. Exp on Streaming Data?

Scanned with OKEN Scanner

...d Data Engineer – GCP
...structor – Srinivas
Interview Questions – Company wise

CELEBRATING
14 YEARS

QualityThought

8. if you have req the data at the fame you may get dup rec at the can avoid that duplicate records?
9. You have dup record in the table how you can delete (Query) row numbers and partition by (name) over? Ex:- ( )
10. I have daily dump of ctv file in GCP how would be bulid a pipeline stored in GCP and BQ may Contain rows files are valid?
11. How can you analyse faulty data?
12. Do you have exposure in python snowflake
13. Aws Redshift
14. What are libraries we used in BQ to access database?

Incedo

1. What is Architecture of BQ?
2. Difference b/w OLAP &OLTP?
3. Spark Architecture?
4. What is Data Qness?
5. Pyspark code for duplicate removal?
6. Difference b/w Show and Take?
7. SQL for Union?
8. Difference b/w union & union all?
9. No.of. occurrence Python code?

Infosys

1. Tell me about yourself, work Life balance.
2. How do you validate data from DB to GCS?
3. How do you load data from local to GCS?
4. Column Mismatch
   Record Mismatch
5. How do you deploy the code using CI/CD Teraform?
6. What is the technical most challenging issue faud?
7. What are the other languages have you worked like python, SQL, Pyspark?

Infosys

1. Why I am not able to operator.
   {bosh –operator.py :101}
   Info –ERROR (gc)
2. How would you import custom modules in Composer?
3. You want Ato mate Multistep operator?
4. Hybrid cloud initiative we have Complex?
5. DAG of Airflow
6. Past data transactions availability?
7. How can we retrieve instance?
8. various deployments in cloud community?

Scanned with OKEN Scanner

9. comparison between AWS & GCP?
10. How setup project in GCP?
11. How to create a bucket in GCP?
12. Which permission required to import data in GCP?
13. How to Configure CBT in cloudshell?

Infosys

1. Views in BQ?
2. Authorized view & Materilized view?
3. Audit table BQ?
4. What is Composer?
5. DAG?
6. How you handle batch job?
7. Diff b/w batch & Streaming Data?
8. How you push files and charges in GitHub Command?
9. Define dictionary in Python?
10. How to iterate keys in dict?

Infosys

1. Explain about your project?
2. How to load data from GCS to BQ?
3. Are you handling the Batch jobs?
4. Trouble shoots in DAG?
5. What is DAG?
6. What is CDC, how you will achieve it?
7. How can you define pandas?
8. Differences between Delete and Truncate in Simplilearn
9. What is Truncate and load in BQ?
10. What do you mean by etl
11. Qhat is authorizwd views in BQ?

Kloud9

1. Best Practies on Pyspark?
2. Out of memory in optimization?
3. Cloud function in GCP?

    --> Cloud data fusion

    --> Data bricks (outside)

    --> github steps (fallow)

4. SQL Code

    --> There is 3 tablees (Dept,Employee&Salary)

    --> Dept (Deptcode,Dept Name,Location)

→ Salary(Empcode,Basic,Additions,Deduction)

-→ Sal emname,Sal,max(Sql))from fn) Where Sql 2

Sal max(Sql)from tn where Sql<(Set Mal (SQL)fromtn))

-→ Empy (Empcode,EmpName,EmpName,Email,PhoneNo,Job,HireDate,Deptcode)

-→ Please use these tables to writea query 3rd Highest Salary.

5. Read CSV file into Dataframe file name Data.CSV.It is having column Age where age is greater than 30?

6. RDD -→ List of integers –what the sum of integers in RDD only even integers.

7. Empty table in Data frame find the arage salary of Empy using spark?

8. Diff b/w sparkcontext and sparksession ?

LTI Mindtree

1. Self Introduction?
2. What is Map Reduce?
3. How you used in Map reduce?
4. Adv and Dis adv of dataproc on Hadoop?
5. Can you tell me differ b/w dataproc and Hadoop cluster?
6. Can you tell me the nameof the operator.which would be roll on the dataproc cluster in production environment?
7. What are the services you known in GCP?
8. What are the different classes in GCP?
9. What is versioning is GCS?
10. What was the architencture of BQ?
11. How you store the data in BQ,How you impentent storage in BQ?
12. What are the challengers you faced in processing dats in BQ?
13. How you done the encryption in BQ?
14. What are window functions?
15. Diff b/w rank and dense rank?
16. What is wild card table?Explain?
17. While using multitrending.how can optimise gcs memory by using multireading [memory optimitation]?
18. how to copy multiple objects from local to GCS?
19. Tell me dictionary having any sequence?
20. diff b/w List & Tuple?
21. What is magical functions in python?
22. What is the use of ---init() the method in python?
23. What is the use of main() main method in python?

LTI Mindtree

1. What are GCP components you worked?
2. If you have 5GB datafile, if you are facing problem to process a file in Airflow?How to resolve this?

Engineer – GCP
Srinivas
Questions – Company wise

CELEBRATING
14 YEARS

QualityThought®

11. Object file class in BQ?
12. Difference b/w Snowflake and fact conselation schema?
13. Star schema and snowflake Schema?
14. SCD type 1,2,3 with example?

Neutrino tech labs

1. What are use cases you using in Pubsub?
2. What are we doing BQ?
3. Where you waiting temponary table?
4. Store Procedure in BQ?
5. Views in Big Query?
6. SQL query to batch data between two dates?
7. Create the dag?
8. What modification you do in predefined dag?
9. What operator did you use in dag?
10. What are the usecase you use in airflow dag?
11. Serverless technology in GCP?
12. Cloud function?
13. How you are triggering cloud function?
14. What API you bulid on cloud function?
15. How you will batch data?
16. How many year of emperience in GCP?
17. What are the usecase used on the python script?

Openreach.co

1. Phython code

   X='aabbadcaa'

   Output=a3b2a1d1c1a2

2. In Python
   Input: [1,1,1,2,3,3,3,4,4,4,4]
   Output: [(1,3),(2,1),(3,3),(4,4)]
3. What is Query data optimization The whole dump the data into the GCS bucket data ?
4. We created table in GCS then how could to transfer that whole data into GCS bucket?
5. Is there import & export option in Big Query?
6. In which Senario we use External table & native table ?
7. Can we use Partition techniques?
8. In what Senario we create External table?
9. How can you resolve Containes killed on request exit code is 137?
10. One file is present in GCS Bucket by using Airflow Compress?
11. In Airflow you need to insert and operate now data?
12. What Sytnax we use in Airflow?
13. Table my numbers cointaxs many numbers is column numbers Including duplicated once,can you write a SQL query to find the biggest?

Cloud Data Engineer – GCP
Instructor – Srinivas
Interview Questions – Company wise

CELEBRATING
14 YEARS

QualityThought

PPT

1. Transformation is cloud on ANDHOOT KUBAI Bigquery?
2. How much size data you're handling in BigQuery?
3. How ingestion process data load from GCS to Bigquery?
4. Where are you using dataflow or dataproc?
5. What usecare is used for dataflow and dataproc?
6. What skillset you are using in project?
7. Apart from streaming what are the other advantages for dataflow and dataproc?
8. In same table we have different set of roles access to rows in Bigquery?
9. Bigquery snapshot?
10. In Bigquery ingertion messedup the data?
11. To see the data two days black on the same table in Bigquery?

Prolify tech

1. Difference b/w List and turple?
2. How do you read date from cv
3. How do you handle errors is python?
4. How do you handle missing values in datasets?
5. Concepts of ETL?
6. Difference b/w ETL and ELT?
7. In which situation use ETL and ELT?
8. Distributed data processing from work in python?
9. How do you use Apache spark?
10. Explain in airflow ,key component in airflow?
11. DAG processing?
12. Task scheduling and execution in airflow?
13. What DB used in you project?
14. DDL commands in SQL?
15. Truncate vs Delete?
16. Where cloud vs Group By?
17. Write simple SQL query using group by?
18. 18.What are joins in SQL?
19. 19.What is use of left join?
20. usr for limit and offset in SQL?
21. Which command used the host limit or offset?
22. Views in SQL?
23. What are acid preparations in SQL?
24. What is Subquery?
25. What is Apche Khafra?
26. Experience in snowflake?

RandomTree

1. Self introduction?
2. Dealing with batch & Streaming files?

3. Dataflow, Custom retransformations?
4. Challenges you faced as dataengineer?
5. Cron job in composer/Airflow 6hr of data is delivered→dag→event driven dag→What is the dag theme?
6. Airflow sensor?
7. Deferable operators→(Postpone executive operators)?
8. Native tables or tables you worked?
9. What kind of unstructed data you are getting text files,images, videos, logfiles, Social media data etc?
10. What exactly log files like system logs or Historical Lags?

Sadia

1. Source System you are getting in GCS?
2. Who are impossible to move file to GCS?
3. Which form of files you working?
4. In which class storage file is kept?
5. How you Processing data from stage to history?
6. Which operator you using in airflow to move GCS to Bigquery?
7. How you communicating in airflow if a particular file is arrived and ready for processing?
8. How are we using cloud functions and airflow together?
9. Datalake vs Datawarehouse?
10. In bigquery how you storing data?
11. On an average how much data you processing?
12. Standard practice storing data in bigquery?
13. What is partioning?
14. Way to retirve a deleted table?
15. BigQuery Architecture?
16. Types of Storage class available?
17. Worked on dataflow?
18. list-1 =["a","c","1",1,2,3,4,5] Seqregate string into one list.Numbers into another list. Explain the code?

Sapient

1. Pipeline in GCP?
2. How you Managing defencies ,sheduling another DAG?
3. How many dags you run daily?
4. Sftp operator?
5. Trigger Dag run operator?
6. How do you do masking in Big Query?
7. Query for encryption and description?
8. SQL-Query:

EmpNo	DeptNo	SAL
7934	10	1300

Cloud Data Engineer – GCP
Instructor – Srinivas
Interview Questions – Company wise

*Quality*Thoug

CELEBRATING
14 YEARS

3. In Excel you have 10 sheets, how to read 5th excel sheet to in the python?
4. Different ways of merge?
5. Difference b/w update & insert and delete & insert in merge?
6. which is a costlies in the above?
7. What is SCD?
8. How do you implement SCD in Big Query?
9. Referential integrity?
10. In airflow if I want skip first 5 records and load file from 6th record?
11. How to load first five(5) file record in file in airflow?
12. If we are facing issue for reading. Table which is taking long time: How do you read the table.
13. On which column we will provide partition?
14. How to ZIP file in Python?
15. How to migrate dataset from on region to another region?

LTI Mindtree

1. Table

   Student   Student name   Table marks

   Get Marks & Dense Rank marks on Student table in one statement?

   Select st.name Total marks.

   Rank () over (order by total marks desc) as rank

   Dense rank () over (order by total marks desc) as d.rank

   From student,

2. Total avg of student?
3. Write a code for prime numbers in python?
4. Write a SQL query for joining 2 tables where sal>5000 and Location is Pune using Left join?
5. Write some code using code?
6. What is dag in Airflow & Explain the Dag?

LTI Mindtree

1. Self Introduction
2. Direct transform from Hive to BQ (How you tray) by using tool?
3. How you migrate data from on premises?
4. Challenges you faced while migrating data from onpremise to cloud?
5. Top 5 services you work with?
6. Explain BQ Architecture?
7. Nested and Repeated fields in BQ?
8. Time travel in BQ?
9. BQ –cost optimization techniques in BQ?
10. Use case in Cloud composer.what type of Scripts you me using?

Scanned with OKEN Scanner